

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2003, 09:15:22 ; Search time 39 Seconds

(without alignments)
2745.944 Million cell updates/sec

Title: US-09-068-377c-1

Sequence: 1 MAAQGFDAFCRCDETAHT.....WTVERNGRGVPGSYLEKL 415

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	2152	100.0	415 11 P97814	P97814 mus musculu
2	1913.5	88.9	416 4 O43586	O43586 homo sapien
3	1877.5	87.2	416 4 O93657	O93657 homo sapien
4	1748	81.2	397 4 O43585	O43585 homo sapien
5	343	15.9	389 5 O8MML2	O8MML2 dictyosteli
6	317.5	14.8	533 5 O8MY10	O8MY10 dictyosteli
7	305.5	14.2	177 4 O96HR6	O96HR6 homo sapien
8	288.5	13.4	494 4 O9VD11	O9VD11 drosophila
9	283	13.2	889 4 O8IW22	O8IW22 homo sapien
10	283	13.2	306 4 O14526	O14526 homo sapien
11	272	12.6	376 4 O96CF5	O96CF5 homo sapien
12	257	11.9	785 11 O8C012	O8C012 mus musculu
13	256	11.0	476 11 O8K285	O8K285 mus musculu
14	237.5	11.0	476 6 O8HXC5	O8HXC5 macaca fasc
15	228	10.6	506 4 O8IV19	O8IV19 homo sapien
16	223	10.4	471 5 O8MMW4	O8MMW4 dictyosteli

17	216	10.0	1333 5 O8INB8	O8INB8 drosophila
18	207.5	9.6	968 5 O9TXS0	O9TXS0 caenorhabdi
19	204.5	9.5	1112 5 O9VHC4	O9VHC4 drosophila
20	199.5	9.3	1011 5 O61639	O61639 drosophila
21	199.5	9.3	1014 5 O8INU2	O8INU2 drosophila
22	199.5	9.3	1094 5 O61618	O61618 drosophila
23	199.5	9.3	1097 5 O9VIF7	O9VIF7 drosophila
24	192	8.9	546 11 O8BTR8	O8BTR8 mus musculu
25	191.5	8.9	547 5 O07839	O07839 echinococcu
26	191.5	8.9	416 5 O99L10	O99L10 mus musculu
27	191	8.9	1734 11 O9J109	O9J109 rattus norv
28	191	8.9	1793 11 O9J1R0	O9J1R0 rattus norv
29	190	8.8	582 4 O98B88	O98B88 homo sapien
30	190	8.7	620 11 O8RS11	O8RS11 rattus norv
31	187.5	8.7	679 4 O96RU3	O96RU3 homo sapien
32	186.5	8.7	547 11 P97531	P97531 rattus norv
33	186	8.6	905 4 O9BX64	O9BX64 homo sapien
34	185.5	8.6	2427 5 O21408	O21408 caenorhabdi
35	183	8.5	378 13 O8AVN0	O8AVN0 xenopus lae
36	183	8.5	951 5 O16647	O16647 caenorhabdi
37	179.5	8.3	433 11 O62418	O62418 mus musculu
38	179.5	8.3	439 4 O96K74	O96K74 homo sapien
39	179.5	8.3	701 4 O9NXX8	O9NXX8 homo sapien
40	179	8.3	740 11 O8K3Y2	O8K3Y2 mus musculu
41	178.5	8.3	471 11 O70419	O70419 rattus norv
42	178	8.3	432 11 O8BH56	O8BH56 mus musculu
43	178	8.3	509 11 O921L6	O921L6 mus musculu
44	178	8.3	509 11 O8BNAS	O8BNAS mus musculu
45	177.5	8.2	426 5 O07840	O07840 echinococcu

ALIGNMENTS

RESULT 1

P97814

PRELIMINARY:

PRT: 415 AA.

AC P97814;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE PEST phosphatase interacting protein.
GN PSTRIP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

RM [1]
RP SEQUENCE FROM N.A.

RA Laaky L.;

RT "PEST Phosphatase Interacting Protein (PSTRIP1).";

RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.

DR EMBL: U87814; AAB48483.1; -

DR HSSP: Q60631; IGBO.

DR MGD: MGI:1321396; Pstrp1.

DR InterPro: IPR001060; Cdc15_fes_CIP4.

DR InterPro: IPR001452; SH3.

DR Pfam: PF00611; FCH; 1.

DR PRINTS: PR00452; SH3DOMAIN.

DR PRODOM: PD000066; SH3; 1.

DR SMART: SM00055; FCH; 1.

DR SMART: SM00326; SH3; 1.

DR PROSITE: PS50133; FCH; 1.

DR PROSITE: PS50002; SH3; 1.

DR SH3 domain.

SO SEQUENCE

Query Match 415 AA; 47590 MM; 16C0329284D2739C CRC64;

Best Local Similarity: 100.0%; Score 2152; DB 11; Length 415;

Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MMAOLQFRAFCRDPDTAHTGTEVLLQRLLDGRKKMKDVEELLRQRAQAEEERYGKELVQI 60
DB 1 MMAOLQFRAFCRDPDTAHTGTEVLLQRLLDGRKKMKDVEELLRQRAQAEEERYGKELVQI 60
QY 61 ARKAGGQTEMNSLRFSFSLKQOOTEVNGSAHIQALALREELSLSEEFREKQKRYE 120
DB 61 ARKAGGQTEMNSLRFSFSLKQOOTEVNGSAHIQALALREELSLSEEFREKQKRYE 120
QY 121 AIMDRVQKSLSYKTKMESKKAYDQCKDADAEQAFERVSANGHOKOVEKSONKAKOC 180
DB 121 AIMDRVQKSLSYKTKMESKKAYDQCKDADAEQAFERVSANGHOKOVEKSONKAKOC 180
QY 181 KESATAEAEVRYRONITOLERARTEMEOEHRITCEAFQLOEFDRLLTLRNALWVHCNOLSM 240
DB 181 KESATAEAEVRYRONITOLERARTEMEOEHRITCEAFQLOEFDRLLTLRNALWVHCNOLSM 240
QY 241 QCVKDELVEEYRLTEGCDVEEDINGFIQSKSTGREGPPAPVYQNYVDREVTPPLIGSPS 300
DB 241 QCVKDELVEEYRLTEGCDVEEDINGFIQSKSTGREGPPAPVYQNYVDREVTPPLIGSPS 300
QY 301 IQPSCGVIRKFSGLHSGSKPTTPSAPASTETLLTPPERNELVYASIEVOATQGNLSSA 360
DB 301 IQPSCGVIRKFSGLHSGSKPTTPSAPASTETLLTPPERNELVYASIEVOATQGNLSSA 360
QY 361 QDYRALYDYTAQNSDELDISAGDILAVILLEGEDEGMWTVERNQGRGFVGSYLEKL 415
DB 361 QDYRALYDYTAQNSDELDISAGDILAVILLEGEDEGMWTVERNQGRGFVGSYLEKL 415

RESULT 2
O43586 PRELIMINARY; PRT; 416 AA.
AC 043586;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE CD2 binding protein 1 long form (Proline-serine-threonine phosphatase
DE interacting protein 1).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99077800; PubMed=9857189;
RA Li J., Nishizawa K., An W., Husey R.E., Lialios F.E., Salgia R.,
RA Sunder-Plassmann R., Reinherz E.L.;
RT "A cdcl5-like adaptor protein (CD2BP1) interacts with the CD2
RT cytoplasmic domain and regulates CD2-triggered adhesion.";
RL EMBO J. 17:7320-7336(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RA Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL: AF038603; AAD11959.1;
DR EMBL: BC008602; AA08602.1;
DR HSSP: P07751; 1BK2.
DR Genew: HGNC:9580; PSTPIP1.
DR InterPro: IPR001060; Cdc15_Fes_CIP4.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00611; FCH_1.
DR Pfam: PF00018; SH3_1.
DR PRINTS: PR00452; SH3DOMAIN.
DR PRODOM: PD000066; SH3_1.
DR SMART: SM00055; FCH_1.
DR SMART: SM00326; SH3_1.
DR PROSITE: PS50133; FCH_1.
DR PROSITE: PS50002; SH3_1.
DR SH3 domain.
SQ SEQUENCE 416 AA; 47591 MW; 97818150B3D5D600 CRC64;

```

```

Query Match 88.9%; Score 1913.5; DB 4; Length 416;
Best Local Similarity 88.0%; Pred. No. 1,1e-118;
Matches 366; Conservative 20; Mismatches 29; Indels 1; Gaps 1;

QY 1 MMAOLQFRAFCRDPDTAHTGTEVLLQRLLDGRKKMKDVEELLRQRAQAEEERYGKELVQI 60
DB 1 MMAOLQFRAFCRDPDTAHTGTEVLLQRLLDGRKKMKDVEELLRQRAQAEEERYGKELVQI 60
QY 61 ARKAGGQTEMNSLRFSFSLKQOOTEVNGSAHIQALALREELSLSEEFREKQKRYE 120
DB 61 ARKAGGQTEMNSLRFSFSLKQOOTEVNGSAHIQALALREELSLSEEFREKQKRYE 120
QY 121 AIMDRVQKSLSYKTKMESKKAYDQCKDADAEQAFERVSANGHOKOVEKSONKAKOC 180
DB 121 AIMDRVQKSLSYKTKMESKKAYDQCKDADAEQAFERVSANGHOKOVEKSONKAKOC 180
QY 181 KESATAEAEVRYRONITOLERARTEMEOEHRITCEAFQLOEFDRLLTLRNALWVHCNOLSM 240
DB 181 KESATAEAEVRYRONITOLERARTEMEOEHRITCEAFQLOEFDRLLTLRNALWVHCNOLSM 240
QY 241 QCVKDELVEEYRLTEGCDVEEDINGFIQSKSTGREGPPAPVYQNYVDREVTPPLIGSPS 300
DB 241 QCVKDELVEEYRLTEGCDVEEDINGFIQSKSTGREGPPAPVYQNYVDREVTPPLIGSPS 300
QY 301 IQPSCGVIRKFSGLHSGSKPTTPSAPASTETLLTPPERNELVYASIEVOATQGNLSSA 359
DB 301 IQPSCGVIRKFSGLHSGSKPTTPSAPASTETLLTPPERNELVYASIEVOATQGNLSSA 359
QY 361 QDYRALYDYTAQNSDELDISAGDILAVILLEGEDEGMWTVERNQGRGFVGSYLEKL 415
DB 361 QDYRALYDYTAQNSDELDISAGDILAVILLEGEDEGMWTVERNQGRGFVGSYLEKL 415

RESULT 3
O95657 PRELIMINARY; PRT; 416 AA.
AC 095657;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE PEST phosphatase interacting protein homolog.
DE H-PIP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wilson L.A., Fields D., Cruz L., Lasky L., Friesen J.,
RA Simnovitch K.A.;
RT "The human homologue of mouse PTP-PIP interactor protein.";
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL: U94778; AAD00762.1;
DR HSSP: O60631; 1GB0.
DR InterPro: IPR001060; Cdc15_Fes_CIP4.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00611; FCH_1.
DR Pfam: PF00018; SH3_1.
DR PRINTS: PR00452; SH3DOMAIN.
DR PRODOM: PD000066; SH3_1.
DR SMART: SM00055; FCH_1.
DR SMART: SM00326; SH3_1.
DR PROSITE: PS50133; FCH_1.
DR PROSITE: PS50002; SH3_1.
DR SH3 domain.
SQ SEQUENCE 416 AA; 47610 MW; 7CED24A2E1A7EEFF CRC64;

Query Match 87.2%; Score 1877.5; DB 4; Length 416;
Best Local Similarity 86.1%; Pred. No. 2,6e-116;
Matches 358; Conservative 23; Mismatches 34; Indels 1; Gaps 1;

QY 1 MMAOLQFRAFCRDPDTAHTGTEVLLQRLLDGRKKMKDVEELLRQRAQAEEERYGKELVQI 60

```


xx New PEST-type protein tyrosine phosphatase interacting polypeptide -
 PT nucleic acids and vectors; for inducing the polymerisation of actin
 PT monomers in eukaryotic cells and identifying antagonists

PS Claim 1; Fig 1; 111pp; English.

xx The present sequence represents murine tyrosine phosphorylated cleavage
 CC furrow-associated protein (PSTPIP), which is a PEST-type protein
 CC tyrosine phosphatase (PTP)-interacting polypeptide. PSTPIP induces the
 CC polymerisation of actin monomers in a eukaryotic cell, by introducing
 CC a vector containing the nucleic acid sequence encoding PSTPIP into the
 CC cell. Assays for identifying (antagonists of PSTPIP comprise contacting
 CC PSTPIP with the agent and monitoring the ability of PSTPIP to induce
 CC actin polymerisation.

xx Sequence 415 AA:

Query Match 100.0%; Score 2152; DB 19; Length 415;
 Best Local Similarity 100.0%; Pred. No. 3e-185;
 Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMAQLQFRDAFWCGRDFAHNGEYVLLQRLDGRKMKCVDELLRQRAQAEERYGKELVQI 60
 DB 1 MMAQLQFRDAFWCGRDFAHNGEYVLLQRLDGRKMKCVDELLRQRAQAEERYGKELVQI 60
 QY 61 ARKAGGQTEMNSLRTSPDSLKQOTENVGSAHIQALALRELRLSLEEFREEROKRKYE 120
 DB 61 ARKAGGQTEMNSLRTSPDSLKQOTENVGSAHIQALALRELRLSLEEFREEROKRKYE 120
 QY 121 AIMDRVOKSKLSLYKKTWESKAYDQKCRDADAEQAFERVSANGHOKQVEKSONKAKOC 180
 DB 121 AIMDRVOKSKLSLYKKTWESKAYDQKCRDADAEQAFERVSANGHOKQVEKSONKAKOC 180
 QY 181 KESATEARVYRONIEOLERARTEWDEHRTTCEAFOLOEFDRLTLRNALVHCHQOLSM 240
 DB 181 KESATEARVYRONIEOLERARTEWDEHRTTCEAFOLOEFDRLTLRNALVHCHQOLSM 240
 QY 241 QCVKDELLEYEVRLLTEGCDVEGDIINGFIQSKSTGREGPAVPYQNYVREYVPLIGSPS 300
 DB 241 QCVKDELLEYEVRLLTEGCDVEGDIINGFIQSKSTGREGPAVPYQNYVREYVPLIGSPS 300
 QY 301 IQPSCGVYKRRSGLLHSGPKTTPSAPASTETLPTPERNELVYASIEVOATQGNLNSSA 360
 DB 301 IQPSCGVYKRRSGLLHSGPKTTPSAPASTETLPTPERNELVYASIEVOATQGNLNSSA 360
 QY 361 QDYRALDYTAQNSDELIDISAGDILAVILLEGEDGMWTVERNRGQGVPGSYLEKL 415
 DB 361 QDYRALDYTAQNSDELIDISAGDILAVILLEGEDGMWTVERNRGQGVPGSYLEKL 415

RESULT 2
 AAB08486 ID AAB08486 standard; Protein: 415 AA.

XX AAB08486:

XX 20-DEC-2000 (first entry)

XX A murine PTP phosphatase interacting protein (PSTPIP).

XX protein tyrosine phosphatase; PTP; phosphatase interacting protein;
 KW PSTPIP; PEST family; protein tyrosine phosphatase; actin monomer;
 KW tissue typing; tumour cell; tumour imaging.

XX Mus sp.

XX US611073-A.

XX 29-AUG-2000.

XX 06-FEB-1998; 98US-0020222.

PR 07-FEB-1997; 97US-0104590.

XX (GENE) GENENTECH INC.

XX Lasky LA:

XX WPI: 2000-586378/55.

DR N-PSDB; AAA64240.

PT Novel PST phosphatase interacting protein useful for inducing
 PT polymerisation of actin monomers and for identifying homologue of PST
 PT phosphatase interacting protein

PS Claim 1; Column 45-48; 48pp; English.

xx The present sequence represents a protein tyrosine phosphatase (PTP)
 CC phosphatase interacting protein (PSTPIP). PSTPIP polypeptides are
 CC bound by and dephosphorylated by the PST family of protein tyrosine
 CC phosphatases. PSTPIP associates with actin. PSTPIP is useful for
 CC inducing the polymerisation of actin monomer in eukaryotic cells by
 CC introducing the polypeptide into the cell. The polypeptide is useful
 CC for identifying and isolating PSTPIP homologues in another mammalian
 CC species, in screening assays to identify antagonists and agonists of
 CC native PSTPIP polypeptide and as molecular weight markers on protein
 CC gels. The PSTPIP nucleic acid is useful for tissue typing of specific
 CC mammalian tissues, for preparing PSTPIP polypeptides by recombinant
 CC techniques, as hybridisation probes for searching cDNA and genomic
 CC libraries for the coding sequence of other PSTPIP analogues and to
 CC isolate homologous genes specifically expressed in tumour cells.
 CC Antagonists of PSTPIP peptide are useful for inhibiting biological
 CC activity of the peptide. Antibodies of PSTPIP are useful to identify
 CC rapidly dividing cells and are used to image tumours comprising such
 CC rapidly dividing cells.

xx Sequence 415 AA:

Query Match 100.0%; Score 2152; DB 21; Length 415;
 Best Local Similarity 100.0%; Pred. No. 3e-185;
 Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMAQLQFRDAFWCGRDFAHNGEYVLLQRLDGRKMKCVDELLRQRAQAEERYGKELVQI 60
 DB 1 MMAQLQFRDAFWCGRDFAHNGEYVLLQRLDGRKMKCVDELLRQRAQAEERYGKELVQI 60
 QY 61 ARKAGGQTEMNSLRTSPDSLKQOTENVGSAHIQALALRELRLSLEEFREEROKRKYE 120
 DB 61 ARKAGGQTEMNSLRTSPDSLKQOTENVGSAHIQALALRELRLSLEEFREEROKRKYE 120
 QY 121 AIMDRVOKSKLSLYKKTWESKAYDQKCRDADAEQAFERVSANGHOKQVEKSONKAKOC 180
 DB 121 AIMDRVOKSKLSLYKKTWESKAYDQKCRDADAEQAFERVSANGHOKQVEKSONKAKOC 180
 QY 181 KESATEARVYRONIEOLERARTEWDEHRTTCEAFOLOEFDRLTLRNALVHCHQOLSM 240
 DB 181 KESATEARVYRONIEOLERARTEWDEHRTTCEAFOLOEFDRLTLRNALVHCHQOLSM 240
 QY 241 QCVKDELLEYEVRLLTEGCDVEGDIINGFIQSKSTGREGPAVPYQNYVREYVPLIGSPS 300
 DB 241 QCVKDELLEYEVRLLTEGCDVEGDIINGFIQSKSTGREGPAVPYQNYVREYVPLIGSPS 300
 QY 301 IQPSCGVYKRRSGLLHSGPKTTPSAPASTETLPTPERNELVYASIEVOATQGNLNSSA 360
 DB 301 IQPSCGVYKRRSGLLHSGPKTTPSAPASTETLPTPERNELVYASIEVOATQGNLNSSA 360
 QY 361 QDYRALDYTAQNSDELIDISAGDILAVILLEGEDGMWTVERNRGQGVPGSYLEKL 415
 DB 361 QDYRALDYTAQNSDELIDISAGDILAVILLEGEDGMWTVERNRGQGVPGSYLEKL 415

RESULT 3

AA81928 ID AA81928 standard; Protein: 415 AA.

QY 361 QDYRALDYTAONSDDELISAGDILAVILEGEDGMMTERNGRGFEVPGSYLEKL 415
DB 361 QDYRALDYTAONSDDELISAGDILAVILEGEDGMMTERNGRGFEVPGSYLEKL 415

	RESULT 4
AAY28291	AAV28291 standard; protein: 416 AA.
-ID	AAV28291 standard; protein: 416 AA.
XX	
AC	AAV28291;
XX	(first entry)
XX	variant of CD2BP1.

AC	29-SEP-1999	(first entry)	variant of CD2BP1.
XX			
DT			
XX			

KW	antigen recogn
KW	
XX	
OS	Homo sapiens.

WO9936534-A1.

av	22-JUL-1995.	98WO-US26699.
PD		
XX	14-DEC-1998;	
DE		

9805-0008420.
13-JAN-1998;
XX
PR
XX

and related

WPI: 1999-444396/37. CD2B1 and
XX identifying agents which modulate signal
XX cytoplasmic tail binding protein,
XX DR
XX

PT 10500: English
transduction, cell adhesion or motility
polynucleotides, useful in
A human cell

Claim 3; page 88-89; 100pp
xx CD2BpL. CD2BpL is a cDNA
xx PS is a variant of CD2BpL. CD2BpL is a cDNA
xx PS can be used to assay for the presence
xx PS modulatory agents identified in
protein is a variant of CD2BpL. CD2BpL is a cDNA
PS modulatory agents identified in

This procedure was used to modulate CD2BPI activity in cell adhesion protein. Anti-CD2BPI antibodies were added to cell culture medium. These antibodies are used to inhibit or enhance signal transduction of CD2BPI in a cell or tissue sample. The antibodies are used to modulate the signal for enhancing signal.

the melanoma can be used as CD2BPI inhibitors are useful for mimics of CD2BPI inhibitors or motility. Mimics of CD2BPI inhibitors or motility. Mimics of CD2BPI inhibitors or motility. Mimics of CD2BPI inhibitors or motility.

transduction, CD281 enhancers have been used to modulate expression of CD281. Enhancement of CD281 is useful for cancer therapy, to augment the efficiency of CD281-related gene delivery.

all classes of individuals and in immunodeficient individuals, inhibition of CD28p1 is useful in cancer cells and in transplantation, CC immune response compromised individuals, include rheumatoid arthritis, CC immunocompromised individuals, and conditions, include hematosis, CC immunocompromised individuals, and conditions, include hematosis, CC

CC useful in such diseases as erythema with tissues
CC diseases. systemic lupus erythematous
CC juvenile diabetes. systemic lupus erythematous
CC adhesion of the recipient's T cells with tissues

	DB 20;	Length 416;
CC	reduce aures-	
CC	auto graft.	
CC		
XX	416 AA;	

Sequence	Score	Indels	Gaps
88.98;	1913.27;		1;
Pred. No. 1e-163;		29;	
88.04;			
20; Mismatches			

[illegible][illegible]

61 ARKAGGOTEMNSLKRIJFDCB ||||| :||||
 | ||||| :||| | |||||
 I PRKGOTEINSLRASFDLKOOMENVSSHIQALATLRREEDRSDEE

61 ANSWERS

Db

—

100


```

OY 61 ARKAGOTENMSLRTSFDLSKOOTENYNSAHIOALALARELSLEEFREKOEORKK-- 118
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 61 ARKAGOTELNSLRASDLSKQOMENYSSHIQIALALRELRLEEFREKOEORKKM 120
OY 119 -----YEAIMRVOKSKLSYKKTKMESKAYADOKCRDADDAEQAFERY 161
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 121 AVFROSDCEMVKSPSWEYAVMDRVOKSKLSYKKTKMESKAYADOKCRDADDAEQAFERI 180
OY 162 SANHOKOVEKSONKAKOCKESATAEARYVRONIEOLERARTMECHRTTCEAFOLQEF 221
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 181 SANHOKOVEKSONKAKOCKESATAEARYVRONIEOLERARTMECHRTTCEAFOLQEF 240
OY 222 DRLTILNALMVHNSQMSQCVKNDDELYEVRLLTEGCDVEGDIINGFTQSKSTGREPPAP 281
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 241 DRLTILNALMVHNSQMSQCVKNDDELYEVRLLTEGCDSDIDADIDSFIOAKSTGTEPP-- 298
OY 282 VPTQNTYDREVTPLIGSPSTQPSGCVTKRPSGLHSGPKTTPSPAASATETLPTTPPRNE 341
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 299 -----RFSGLHSGPKTTPSPAASATETLPTTPPRNE 330
OY 342 LVYASIEVOATQGNLSSADYRALTAYTAONSDLELISAGDILAVLLEGDDGWTVERN 401
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 331 GVTYALAVQELIQNPASPADYRALTAYTAONSDLELISAGD-----ILLEGDDGWTVERN 386
OY 402 GORGFVPGSYLEKL 415
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 387 GORGFVPGSYLEKL 400

```

RESULT 8
ID AAM39877 standard; Protein: 334 AA.
XX AAM39877:
AC AAM39877:
DT 22-OCT-2001 (first entry)
DE Human polypeptide SEQ ID NO 3022.
XX

Human: noctropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
XX WO200153312-A1.
XX
XX PD 26-JUL-2001.
XX
XX PF 26-DEC-2000; 2000MO-US34263.
XX
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 29-NOV-2000; 2000US-0727344.
XX
XX PA (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
XX
XX DR N-PSDB; AA159033.
XX

```

PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 4; SEQ ID NO 3022; 10078bp; English.
XX
XX CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM2213) with noctropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemia and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 334 AA;

```

Query Match 34.2%; Score 736; DB 22; Length 334;
Best Local Similarity 44.2%; Pred. No. 1.2e-57;
Matches 144; Conservative 64; Mismatches 112; Indels 6; Gaps 3;

```

OY 2 MAQLQFADAFWCDFTAHTGYEVLLQRLDGRKCKDVLELLQROAERKELYQTA 61
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1 MTRSLFEGNWSADISTIGYDNIQHLNGRKCKEFEDFLKRAAIEERYKDLNLS 60
OY 62 RKAG-GOTENMSLRTSFDLSKOOTENYNSAHIOALALARELSLEEFREKOEORRKY 120
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 61 RKRPCGSEINTLKRALFEVKQOYDVAQCHIOALOSLREAKKMEFFRKQLOKKTE 120
OY 121 AINDRVOKSKLSYKKTKMESKAYADOKCRDADDAEQAFERYVNSANHOKOVEKSONKAKOC 180
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 121 LINDAIHKOKSLQFKTKMDAKKNVEQKCRDDEAGVNSANLVNPKQOEKLFVKATS 180
OY 181 KESATAEARYVRONIEOLERARTMECHRTTCEAFOLQEFDLTILNALMVHNSQMS 240
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 181 KTVESDSKAYMHIGILDVREMOSEHAKAEAEAOECERINFEKNALMVHNSQMS 240
OY 241 QCVKNDDELYEVRLLTEGCDVEGDIINGFTQSKSTGREPPAPVRYONYDREVTPLIGSPS 300
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 241 QCVTSDEMIVQVKSLEMCISIQRDIEYVQKRTGQIPAPINENFTSSQKNAVAGKA 300
OY 301 IQPSCGVYIKRPSGLHSGPKTTPSPA 326
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 301 TGFN---LARRGPL--PIPKSSPDDP 321

```

RESULT 9
ID AAB64415 standard; Protein: 334 AA.
XX AAB64415:
AC AAB64415:
DT 22-MAR-2001 (first entry)
DE Amino acid sequence of human intracellular signaling molecule INTR447.
XX
XX DE Human: intracellular signaling molecule; INTRA: immunosuppressant;
KW cytoskeletal; neuroprotective; noctropic; antiarteriosclerotic; cancer;
KW antiinflammatory; anti-HIV; neuroleptic; antibacterial; antifungal;
KW antiviral; antiparasitic; antihelminthic; antiparkinsonian; AIDS;
KW cell proliferative disorder; arteriosclerosis autoimmune; epilepsy;
KW inflammatory disorder; Addison's disease; gastrointestinal disorder;
KW neurological disorder; Parkinson's disease; Creutzfeldt-Jakob disease;
KW mental disorder; schizophrenia; anxiety.
XX
XX OS Homo sapiens.
XX

XX MO200077040-A2.
XX 21-DEC-2000.
XX 16-JUN-2000; 2000WO-US16636.
XX 16-JUN-1999; 99US-0139566.
XX 17-AUG-1999; 99US-0149640.
XX 09-NOV-1999; 99US-0164417.
XX (INCY-) INCYTE GENOMICS INC.
XX Yve H, Tang YF, Hillman JL, Lal P, Bandman O, Baughn MR;
XX Azimzal Y, Yang J, Reddy R, Lu DM;
XX WPI; 2001-025334/03.
XX N-PSDB; AAF32684.
XX New human intracellular signaling molecules, useful for the diagnosis,
XX prevention and treatment of cell proliferative, autoimmune,
XX inflammatory, neurological, gastrointestinal, reproductive and
XX developmental disorders -
XX
XX Claim 5: Page 156-157; 192pp; English.
XX
XX Sequences AAF32638 - AAF32689 represent cDNA encoding human
XX intracellular signaling molecules INTRAI - INTRA52, represented in
XX AAB64369 - AAB64420. Modulators of the intracellular signaling molecules
XX of the invention exhibit immunosuppressive, cytostatic, neuroprotective;
XX neurologic; antiarteriosclerotic; antiinflammatory; anti-HIV;
XX CC antihelminthic; and antiparkinsonian activity. INTRA polypeptides their
XX agonists and antagonists are useful for the treatment of a condition
XX associated with decreased or increased expression of functional INTRA.
XX Disorders associated with abnormal INTRA expression or activity include
XX cell proliferative disorders e.g. arteriosclerosis and cancers;
XX autoimmune or inflammatory disorders e.g. Addison's disease and acquired
XX immunodeficiency syndrome (AIDS); viral, bacterial, fungal, parasitic,
XX protozoal and helminthic infections; gastrointestinal disorders e.g.
XX dysphagia and irritable bowel syndrome; neurological disorders e.g.
XX epilepsy and Parkinson's disease; prion diseases e.g. Creutzfeldt-Jakob
XX disease and mental disorders e.g. anxiety, schizophrenia and Tourette's
XX disorder. Antibodies immuno specific for the INTRA proteins may also be
XX useful in the diagnosis of the above disorders.
XX
XX Sequence 334 AA:
SQ
Query Match 34.2%; Score 736; DB 22; Length 334;
Best local Similarity 44.2%; Pred. No. 1.2e-57;
Matches 144; Conservative 64; Mismatches 112; Indels 6; Gaps 3;
QY 2 MAQLQFRDAFWCRDFAHTGYEVLQRLDGRKKCKVBEELRQRAQAEERYGKELVOIA 61
DB 1 MFSLSLFGNWSADILSTIGYDNIQLHNGRKCKEFEPFLERAIERYGKDLNLS 60
QY 62 RKAG-GOTENNSLRTPSDLSKOOTENYNSAHIOALALRELSLEFFEROKRQRYE 120
DB 61 RKPPCGOSELNTRKRLAEVFKQVDNAQCHIOALQSLREARKMEFERKOKIQRKTE 120
QY 121 AINDRYOKSKLSLYKTMESKKAAYDOKCRDADAEQAFEEVNSANGHOKOVEKSONKAKOC 180
DB 122 LINDALHKKOSLOPKFTMOAKKNYEOKCRDKDEAQVANSANLVNPKQOEKLEFKLATS 180
QY 181 KESATTAERYRONIOLEFARFWEDEOHTTCFAFOLQFDRITLIRNALWQCNLSM 240
DB 181 KTVAVESDKAYMHGIGLTKVREMOSEHFKACEAFQACERINFRNALMLHVNLSQ 240
QY 241 QCVKDELVEEVLTLEGQVEDINGEIOSKSGRPPAPVPYQNYDYDEVPLISPS 300
DB 241 QCVTSDEMTQVRSKLEMSIQRDIEYVQNRKTGQIPAPIMVENEYSSOKNAVPAGKA 300
QY 301 IOPSCVIRKFRSGILHSGPKTPPSAP 326

DB 301 TGFN---LARRGPL-PIPKSSPDP 321
RESULT 10
AAM41663
ID AAM41663 standard; Protein: 365 AA.
XX AAM41663;
XX 22-OCT-2001 (first entry)
XX Human polypeptide SEQ ID NO 6594.
XX Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokine; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia.
XX Homo sapiens.
XX MO200153312-A1.
XX 26-JUL-2001.
XX 26-DEC-2000; 2000WO-US34263.
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0634450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
XX N-PSDB; AAI60819.
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Example 2: SEQ ID NO 6594; 10078bp; English.
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AAM38642-AAM42213) with neurotropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokine activity, haemostatic,
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX Sequence 365 AA:
SQ
Query Match 34.2%; Score 736; DB 22; Length 365;
Best local Similarity 44.2%; Pred. No. 1.4e-57;

PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209457.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 11-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236337.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 02-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
PA
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465460/50.
DR N-PSDB; AAS27373.
XX
XX Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders
XX
XX
PS Claim 1; SEQ ID NO 1021; 880pp; English.
XX


```

XX      Sequence      486 AA:
SQ      Query Match      13.2%; Score 284.5; DB 24; Length 486;
          Best Local Similarity 21.1%; Pred. No. 1.2e-16;
          Matches 102; Conservative 92; Mismatches 200; Indels 89; Gaps 14;

QY      9 DAFWCRDFTAHNG-YEVLRLDGRKMKDVEELLRQRAQAEERYGKELVQIARKAGQ 67
          16 DSFW-----EYGNKRYKRYKRIDDGHRICSDLMNGLHERARIEKAYAOQLTEMARWRQL 69
QY      68 TE---MNSLRTSEDSLKQOTENVGSAHIQALALREELRLSEFREROKE-----114
          70 VEKGPQYGVYKEMWAFMSEAEYVSELHLEVKASLMD--DEKIKNMQKEAFHKOMMG 127
QY      115 --QRKKYEAIMDRVQSKSLYKKTMSKKAYDOKCRDADA--EOAFERYVANGHOKOV 170
          128 FKETKEAEDGFRKAKQPMKAKLKEVEAKKAHHAACKKEKLAISRANSKADPSFNPEDL 187
QY      171 EKSQNKAKCKESATEAEERYVRQNTIEQLERARTWEOEHRTTCEAFQLOEFDRLTLRNA 230
          188 KKLQDKIEKCKQDVLKTKKYEKSLKELDQGPQYVEMNEQVFEQCOQFEKKRLRFFREV 247
QY      231 LMVHCNQLSMQCYKD-DELYEVRRLTEGCDYEGDINGFIQSKSTGREPPAPYPYQNYD 289
          248 LLEVQKHNLNSNAGYKALYHDEOSIRADAVEDLRWFRANHGPMAMNP-----QTE 302
QY      290 REVTPFLIGSPSIOPS-----CGVIRFSGSLHSGPKTT--PSAPAST-----330
          303 EMSADLIRLRSREKKKATDGLTGINQDQFLPSKPSSTLNVPSPNPAQSAQSOSYN 362
QY      331 -----ETLPTPERNELVYASIE-----VOATQGNLSSAQD---362
          363 PREDEDDTGSTVSEKEDIKANKVSYEKTQSYPTDWSDESNNPFSSTDANGDSNPFDD 422
QY      363 -----YRALYDYTAQNSDELISAGDILAVILEGEDGQWTVR--NGQRGFVPGSYL 412
          423 ATSGTEVRYRALYDYGQEHDELSPKAGDELTKMEDDEDEQGWCKGRLDNGQVGLYPANYV 482
QY      413 EKL 415
          483 EAI 485
DB

```

Search completed: July 28, 2003, 09:16:17
 Job time : 47 secs


```

Db      369 KKSFLSKFKLTSRSPFNPNVGNANPAOLSSPRNDSPILTSADDECKMILS--LOEERPQNPT 426
QY      323 PSAPAASTETTLTPPERNEL----- 342
      427 PAAPCAFPNSNTLPRYNELGSLPSPNSVSTEDSRPNVNTPSRQOIOEFGSVLOMEM 486
QY      343 ----VYAS-----LEVATOGNL----- 356
      487 RAVSPVYDSRKNGSSSFTLRKSRSPKRPSSLSQNASRLPRSLTPGNLEPNYDFGVYRD 546
QY      357 ----NNSAOD----- 362
Db      547 PASGAPPDDEPYTDRODSFVDDTINTKATGNTSNRLSLPAYPTDGGDTSIDNPJTSDGO 606
QY      363 -----YRALYDYTAONSDELDISAGDIILAVILEGEDGW-----TYERNQORGFVPSSYLE 413
      607 RILGVSALYDYDAIPEEISPRKGDITAVILKLYEDGWMGEFVVGEDDHNRRQGFPSNFR 666
QY      414 KL 415
      667 EI 668
Db

```

RESULT 2

phosphoprotein; fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Aug-2002
C:Accession: J38127; A57087; S54119
R:Badcock, K.; Churcher, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, May 1997
A:Reference number: Z21772
A:Accession: J38127
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-927 <BAD>
A:Cross-references: EMBL:Z95334; PIDN:CAB08599.2; GSPDB:GN00066; SPDB:SPAC20G8.05C
A:Experimental source: strain 972h-; cosmid C20G8
R:Fankhauser, C.; Raymond, A.; Cerutti, L.; Utzig, S.; Hofmann, K.; Simanis, V.
Cell 82, 435-444, 1995
A:Title: The Schizosaccharomyces pombe cdc15 gene is a key element in the reorganization
A:Reference number: A57087; MUID:95360987; PMID:7634333
A:Accession: A57087
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 'MLTKSLQ', 28-927 <FAN>
A:Cross-references: GB:X86179
R:Fankhauser, C.; Raymond, A.; Cerutti, L.; Simanis, V.
submitted to the EMBL Data Library, April 1995
A:Description: The cdc15 gene is a key element in F-actin reorganisation at mitosis.
A:Reference number: S54119
A:Accession: S54119
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 'MLTKSLQ', 28-870, 'GKFIKEKL' <FAZ>
A:Cross-references: EMBL:X86179
C:Genetics:
A:Gene: SPDB:SPAC20G8.05c; cdc15
A:Map position: 1
A:Introns: 27/3: 58/2: 871/1
C:Superfamily: fission yeast scd2 protein; SH3 homology
C:Keywords: mitosis; phosphoprotein
;873-924/Domain: SH3 homology <SH3>

```

Query Match      16.1%  Score 345.5  DB 2:  length 927:
Best Local Similarity 25.9%  Pred. No. 3.1e-13:
Matches 96:  Conservative 74:  Mismatches 169:  Indels 31:  Gaps 8

OY      3 AQLDFRAFCRDTAHTGTVLLQRLLDGRMKDVLELLQRAQAERYGKELVQIAR 62
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      18 ISSVFRNFPMQSE--DAGMDALSRKSSLSVESIDIEFAKRAISIEEYVSKIQEAA 74

```

```
OY      63 KAGQOEMNSLRTPSLSKQOTENNGSAIQALALREELAS-LEEFERKQEDORKRYEA 1271
Db      75 SSADIDEVGSLTNILSMRTETGSMAKAHEEVSQIINTELNNKIREDYIDQTEQQVAAAN 1340
OY      122 IMRVOKSRKSLTKKTMESEKKAYVDCKCRADADAEGAIFRVRANGHOVEQSOKAKOCK 181
Db      135 AIEELLQJKTALTIDISEKKDAYESCINKLNYMOTKKMG---RELDKYNLKIRIOAA 1900
OY      182 ESATSEARVYRONIEOLERAPTEMQEJHRTTCFAFOLEPRLITLNALVAHCNOLSMQ 2411
Db      191 LAWKKNDAEYREINELLVTYREINDRMTGYODAFQHIEERYLFELKTMNMATANIISTA 250
OY      242 CVKDDELYEEVRLTLEGCEVEDINGFIQSKSTGEHPAPVYONYNDRE-----VTPL 2955
Db      251 CVKODESCERIKLTLENTNIDEDIOMQINEGTGITP-PLPEENDYFKKEUNGUNYDIQL 3099
OY      296 IGSPTSIQPSCGVIKRRSGILHGSPPTTSAPAESTETLT-----PT-----PERN 3404
Db      310 ISKAPSYPPSSSHPSASIASLSSPRSAFPRXTS-ETVSSSEVWSSPPTSPLHSVPKPVSN 3688
OY      341 ELVYASIEVO 350
Db      369 EOYEQVTEVE 378
```

RESULT 3

hypothetical protein KIAA0290 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 05-Nov-1999
C:Accession: T00039
R:Ohara, O.; Nagase, T.; Ishikawa, K.; Nakajima, D.; Ohura, M.; Seki, N.; Nomura, N
submitted to the EMBL Data Library, August 1997
A:Description: Prediction of the coding sequences of unidentified human genes.
A:Reference number: Z14073
A:Accession: T00039
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-906 <OHAS>
A:Cross-references: EMBL:AB006628; NID:d1170681; PID:BAA22959.1; PID:d1023835
A:Experimental source: brain
C:Genetics:
A:Note: KIAA0290

[illegible]


```

Db 493 KLQDIRCR-----LTTQREHEIESTNKSRELRLAETHLQOOLQESQOLLGKMIPEKOSLID 548
Qy 134 -----YKKTMESSKAYDOKRODADDAEOAFERSA-----NGHOKOV 170
Db 549 QLKQVOQNSLRHDSLLTKRALETRKIGRQDLDDLEVEKTRAKLOEIDVFNNQLEL 608
Qy 171 EKSQKAKOCKESATEARVYRONIEOLERANTEWEOHRTTCEAFOLQDFRLTIL-RN 229
Db 609 RELYKQOFQKQODETEKIKOKELE-----RKTSLELDKLKEEDKRRMLQD 655
Qy 230 ALMV-HCNQLSMOCVKDELYEVLTLGCDVEGDINGFIQSKSTG--REPPAP----- 281
Db 656 KLMDRVKQOEERKYKQDEEKEKREVSQKCEVEKKPE--IQEKNKPPHQPPPEPKLGG 713
Qy 282 -VPYQNYDYREVTPLIGSPSIQPSGCVIKRFSGLHGSPTKTPSPAPASTETLPTPEPN 340
Db 714 QIPMWN-----TEKAP-----LTINQDV 732
Qy 341 ELVYASIEVQATQGNLMSAADYRALDYTAQNSDELDISAGDILAV--ILEGEDGWTV 398
Db 733 KVVY-----YRALYPFDAKSHDEITIEPGDILMDESQTEGPGWLGG 774
Qy 399 ERNGQGFVPGSYLEKL 415
Db 775 ELKGTGWFPANYAERM 791

```

RESULT 7

```

B48580
antigen EG13 - tapeworm (Echinococcus granulosus)
C:Species: Echinococcus granulosus
C:Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 23-Sep-2002
C:Accession: B48580
R:Frosch, P.M.; Geier, C.; Kaup, F.J.; Muller, A.; Frosch, M.
Mol. Biochem. Parasitol. 58, 301-310, 1993
A:Title: Molecular cloning of an echinococcal microtichal antigen immunoreactive in Ech
A:Reference number: A48580; MUID:93241221; PMID:8479454
A:Accession: B48580
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-416 <FRO>
A:Cross-references: GB:M9564; NID:q304850; PID:q304851
A:Note: sequence extracted from NCBI backbone (NCBI:130072)
C:Superfamily: echinococcal microtichal antigen EM13; SH3 homology
F:368-416/Domain: SH3 homology <SH3>

```

Query Match 8.9%; Score 191.5; DB 2; Length 416;

Best Local Similarity 21.2%; Pred. No. 0.00021;

Matches 94; Conservative 68; Mismatches 173; Indels 109; Gaps 17;

```

Qy 42 LLROQAQAEERYKELVQIARKAGGOTEMNSLRTSFD-----SLKQOTENVSAMI 92
Db 1 MIQERADIEKAYASNL-----RKFAARLEM-FLRTGVEGTATNIIISGLAKAEDAEIHS 55
Qy 93 QALALREELR-----SLEEFRE--ROKEORRKYEAINDYQKSLSYKKTMESSK 142
Db 56 NIAAGLIPYOLGIKNQRENFHKSISTSIKEYKNFSEFNQKTYKHYKKNVRCK 115
Qy 143 AYDQCRADADAEAFEVNSANGHQYKESQNAKOCKESATEMERVYRONIEOLEPAR 202
Db 116 EYFNACK-----TVRSLOYOVQNAKNEFFGPEQOARQKGLRMEKRLKRG 163
Qy 203 TEMEOEHRTTCEAF-----OLOEFDRLTLIRNALMVHNCNLSMOCV 243
Db 164 MEERKTRAYEEALSSLDVTPRIEDMTQVFNKAQAFERERT-----YVEEQALQMAV 219
Qy 244 KD-----DELYEVLTLGCDVEGDINGFIQSKSTGTPPAPVPPYQNYDYREVTPLI 296
Db 220 LDISAKPMLSQIFVELRETVAKVADADLKKMSLAYGVDMAPNFPV-FQEV----- 269
Qy 297 GSPSIQ-----PSCGV-----IKRFSGLHGS--PKTTPSPAPASTETLPTP 337
Db 270 -SPMSALGKKGRSALADGSSGVTLNLSTKFTSPDRGPIPGTTDSCSNISTSPVHTTD 328

```

```

Qy 338 ERNELVYASIEVQATQGNLNSA--QDY-----RALYDYTAQNSDELDISAGDIL 385
Db 329 YTSYNGAAVAASVSEKQVVEDTPPEYDPFDVDRGCVPIRALYDYVGVADSLSNSGDLF 388
Qy 386 AVIL-EGEDGWTVERNQGRGFPV 408
Db 389 EKLDEDEQGWCKGRKDGVLGP 412

```

RESULT 8

hypoetical protein K10B3.10 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000

C:Accession: T16613

R:Galtung, S.

submitted to the EMBL Data Library, February 1996

A:Description: The sequence of C. elegans cosmid K10B3.

A:Reference number: Z18546

A:Accession: T16613

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2427 <GAT>

A:Cross-references: EMBL:U49941; NID:q1206038; PID:q1206048; PIDN:AAB53876.1; GSPDB:C

A:Experimental source: strain Bristol N2; clone K10B3

A:Genetics:

A:Gene: CESP:K10B3.10

A:Map position: X

A:Insertions: 68/2; 96/3; 138/3; 665/3; 1052/3; 1104/3; 1173/2; 1995/2; 2187/3; 2235/2;

C:Superfamily: spectrin alpha chain; calmodulin repeat homology; SH3 homology; spectr

C:Keywords: EF hand

F:990-1037/Domain: SH3 homology <SH3>

F:2050-2155/Domain: spectrin/dystrophin repeat homology <SPR>

F:2277-2309/Domain: calmodulin repeat homology <EFH>

Query Match 8.6%; Score 185.5; DB 2; Length 2427;

Best Local Similarity 23.5%; Pred. No. 0.0043;

Matches 95; Conservative 67; Mismatches 149; Indels 93; Gaps 21;

```

Qy 37 KDVEELR--QRAQAEERYKELVQIARKAGGOTEMNSLRTSFDLSLKQOTENVSAMITOL 94
Db 705 EDVAMWLELSEGGVASEDYCGDLVSYONL--QKKIGLIESDFNAHNDRVYDGIK---OL 757
Qy 95 ALALREELRSLSEER-----ERKEORRKYEAINDYQKSLSYKKTMESSKAYDQKCR 149
Db 758 AQQFOEE-----EHNAPYVRKQENLQORINALKDPLEKRR---KRLGESHOQ-NOLFR 808
Qy 150 DADDAEQAFERVSANGHQYKESQNAKOCKESATEMERVYRONIEOLEPAR 200
Db 809 DIED-ELAMIR-----EKEQVAGSTNGRDLIGVQNLIKKQALIAELAHNDSQIESVSS 862
Qy 201 ARTEWEOHRTTCEAFOLQEF-DRLTLIRNALMVHNCNLSMOCVKDELYEVLTLGEC 259
Db 863 AANMIOOGH-----FLAPEIRDKLAQLRD-----NWRILKSK---AEKRRT--- 901
Qy 260 DVEGDINGFIQSKSTGTPPAPVPPYQNYDYREVTPLIGSPSIQPSGCVIKRFSGLHGS 319
Db 902 DLDDSLQAH-QYLSDANEADA-----WMSEKEPIVGSFD-----YGKD 938
Qy 320 KTTSPSAPAASETTLTPPERNELVYASIEVQATQ-----GNLNSAADYRALYDYTA 371
Db 939 EDSAEALKKHRLALLSLEAFKGTIEDLRQASQCKQOEOPMQLGDC--VIALYIOE 996
Qy 372 QNSDELDISAGDILAVILSEEDGWTVERNQGRGFPVGSYLEKL 415
Db 997 KSPREVSMMKGDVLTLLNASRDMWKEVANDRQGFVAAYVKRI 1040

```

RESULT 9

protein F56D12.6 (imported) - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: E88042
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A>Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; MUID:99069613; PMID:9851916
A>Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_elegans/
A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: E88042
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-951 <STO>
A:Cross-references: GB:chr_II; PIDN:AB66120.1; PID:q2315677; GSPDB:GN00020; CESP:F56D12.
C:Genetics:
A:Gene: F56D12.6
A:Map position: 2

Query Match 8.5%; Score 183; DB 2; Length 951;
Best Local Similarity 18.9%; Pred. No. 0.0019;
Matches 63; Conservative 83; Mismatches 151; Indels 36; Gaps 10;

QY 23 EVLLQRLDGRKMCVDVDELLRQRAQAEERYKELVQARAGGQTEMNSLFTSFDL-- 80
DB 4 KVLQENLKKSEYVAEQFVKERLSVEDEYKAINRSVKNVSHYIUNGSS--SIDAMWL 60
QY 81 --KOOTENWGSAAHQLALALREELRSLSEEFREROKEORRKY-----EAIMDROYKSKLS 132
DB 61 LTKGTMELMALHVLVANNLQDLSEVLYKEDVNRKRELKQPOVAELV-RLMOTTTTC 119
QY 133 LYKTMESKAYDOKCRDADAEQAFERYVANGHOKOVEKSONKAKOCKESATEAERYVR 192
DB 120 LQK-----AKETVYHNCQLEKVKK-----ETVNVNKEISKVELKVARAREE-----YK 163
QY 193 QNIOLEARFEMWEHHTTCEAFQLOEFDRL--LTLRLALMVHONQLSMOCVKDELYE 250
DB 164 SYVDYELVREDPEFKMSDSCKK--QTFDRSLVSIQOFLFANHSSTEMSSASHQVAE 221
QY 251 EVRLTEGCDVEGDNFTQSKSTGEPPAPYONYNDREYTPILGSPISQGVIKR 310
DB 222 QFKSIOHLMNDEPFRKRVKSTKSTKPRVLFEESENGQLPASSSMNLNLPVADVDVI 281
QY 311 FSGLLHSGPKTPPSAPASTETLPTPRNELV 343
DB 282 MSG--NSMPSSCSSGILQDQAPRPHPTVDLL 312

RESULT 10
A48580
echinococcal microtichal antigen EM13 - tapeworm (Echinococcus multilocularis)
C:Species: Echinococcus multilocularis
C:Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 23-Sep-2002
C:Accession: A48580
R:Frosch, P.M.; Geier, C.; Kaup, F.J.; Muller, A.; Frosch, M.
Mol. Biochem. Parasitol. 58, 301-310, 1993
A>Title: Molecular cloning of an echinococcal microtichal antigen immunoreactive in Ech
A:Reference number: A48580; MUID:93241221; PMID:8479454
A:Accession: A48580
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-426 <FRO>
A:Cross-references: GB:M96565; NID:q304852; PID:q304853
A:Experimental source: larva
A>Note: sequence extracted from NCBI backbone (NCBIN:130067, NCBI:P:130066)
C:Superfamily: echinococcal microtichal antigen EM13; SH3 homology
F:378-426/Domain: SH3 homology <SH3>

Query Match 8.2%; Score 177.5; DB 2; Length 426;
Best Local Similarity 21.2%; Pred. No. 0.0015;
Matches 97; Conservative 66; Mismatches 168; Indels 127; Gaps 19;

QY 42 LLRQRAQAEERYKELVQARAGGQTEMNSLFTSFDL-----SLKQOTENWGSAAH 92
DB 1 MIOERADIEKAYASNL---RKFARLEM-FLRTGVEYGTATNLTSLGLAEADNAELHS 55

QY 93 QALALREELR-----SLEEFRE-----ROKEORRKYEAIMDROKSKLSLYKTMESK 142
DB 56 NIAGILNPVQIGIKMQRNFENFKSSISTSIKEKYNDFEFENAOXTWKHYNNVNRCK 115
QY 143 AYDOKCRDADAEQAFERYVANGHOKOVEKSONKAKOCKESATEAERYVRQNTLEERAR 202
DB 116 EYVHACK-----TVRSLOVQOVAKANPEPTGPQRLKIDKILKKGIMEEKTR 163
QY 203 TEWEQ-----EHRTTCEAF-QLQEDRLTLRLNALMVHONQLSMOCVKD--- 245
DB 164 KAEELALSSLDVTPRYIEDMT--QVFNKQAFERERI-----TYFEQALQMOEVLDISA 217
QY 246 ----DLTYEVRLLTEGCDVEGDNFQSKSGRPPAPYONYNDREYTPILGSPSI 301
DB 218 KPNLSQIFGLRTTAKVDADADLKRMSLAVGVDMANPFPV-FQET-----SPEM 266
QY 302 QPSCGVYKRSGLLHGS-----PKTPPSAPAAST----- 330
DB 267 --SALGKGRSALADGSSGVTLTSLKTLTSPDRGPIPTGDSNISTSPVHTTAYS 324
QY 331 -----ELTPPERNELVYASIEVQATQGNLNSA--QDY-----RALDYTA 371
DB 325 NSYDHSGEATPBDYTSVNGAAASKEKQVEDPYPYDFVDGPRGVPRLADYVG 384
QY 372 QNSEDLDISAGDILAVIL--EGEDGMWTVERNRGGRFVP 408
DB 385 VEADELTFNSGDLFEKLEDEDEQCKGRKRGVGLYP 422

RESULT 11
T39801
hypothetical sh3-containing protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T39801
R:McDougal1, R.C.; Rajandream, M.A.; Barrell, B.G.; Skelton, J.; Churcher, C.M.
submitted to the EMBL Data Library, July 1999
A:Reference number: 221880
A:Accession: T39801
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-501 <MCD>
A:Cross-references: EMBL:AL109731; PIDN:CA852037.1; GSPDB:GN00067; SPDB:SPBC19C2.10
A:Experimental source: strain 972h-; cosmid c19C2
C:Genetics:
A:Gene: SPDB:SPBC19C2.10
A:Map position: 2
A:Introns: 196/3

Query Match 8.2%; Score 177; DB 2; Length 501;
Best Local Similarity 22.1%; Pred. No. 0.002;
Matches 94; Conservative 76; Mismatches 137; Indels 118; Gaps 21;

QY 53 YGKEIVQIRKAGGQTE-MNSLRTSPDSLKQNTENGSAHIQALALREELRSLSEEPRE 111
DB 100 YGKSMEIGHL---QEEFMDYLNST-----LANLESLA---EFKALDKREK 142
QY 112 QKEORRKYAIMDROVOKSKLSLYKTMESKAYDOKCRDADAEQAFERYVANGHOKOVE 171
DB 143 MENRRLVFPALSTRIKAKKE-----ESKLEEDLRNARAK-----YESLE 183
QY 172 KSONKAKOCKESATEAERYVRQNTLEERARKTWEDHRTGCAFOLEF-----DRITIL 227
DB 184 EFERDVOLKE--LEPDRV--ENVVRLQMQIRFHOKSIDLLGLGLEMGFSKKRDVYNI 239
QY 228 R--NALMVHONQLSMOCV-----KDEEL-----YEVRLLTEGCDVEGDI-N 266
DB 240 KRTYSARSLPSNISTSNVATTPSNTLTFEDDTLAKADSSMEYHSPTNTLPSYHTEADLN 299
QY 267 GFTQSKSTGEPPAPYONYNDREYTPL--IGSPSIQ--PSCGVYKRSGLLHSGPK 320
DB 300 SSIASSNRQHT-----EDVNNKDVSDAQNLSIGQSAVDLTTPSSPIPPIR-----HTKPK 348

QY 321 -----TTPSAPAA--STETLTPTPERNELVYASIEV-----QATQGN 355
Db 349 LSTOSTVVKRPSALOSIEDIDLSFKQPELSASSAELEKLSQCVSPSSNTSDAPSK 408
QY 336 LN-----SSADYRALYDYTAONSDELISAGDILAVILEGEGDMMTVERNQORG 405
Db 409 LNRYSYSPLASISSKRVKMYKSFEPETENELTKLKGDLVLVLRKIDEGMWGKLGEDG 468
QY 406 FVPGS 410
Db 469 VFTGN 473
RESULT 12
A47106
myosin heavy chain ID - slime mold (Dictyostelium discoideum)
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Dictyostelium discoideum
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Apr-2002
C:Accession: A47106; B47106
R:Jung, G.; Fukui, Y.; Martin, B.; Hammer III, J.A.
J. Biol. Chem. 268, 14981-14990, 1993
A:Title: Sequence, expression pattern, intracellular localization, and targeted disruption
A:Reference number: A47106; MUID:93315473; PMID:8325874
A:Contents: AX3
A:Accession: A47106
A:Molecule type: DNA
A:Residues: 1-782 <JUN>
A:Note: sequence extracted from NCBI backbone (NCBIN:134989, NCBIIP:134990)
A:Accession: B47106
A:Molecule type: mRNA
A:Residues: 377-1113 <JUN2>
A:Note: sequence extracted from NCBI backbone (NCBIN:134989, NCBIIP:134990)
C:Genetics:
A:Introns: 37/3; 214/1
C:Superfamily: protozoan myosin heavy chain IB; myosin motor domain homology; SH3 homolo
C:Keywords: actin binding; ATP; hydrolase; nucleotide binding; P-loop; tandem repeat
F:10-678/Domain: myosin motor domain homology <MOT>
F:101-108/Region: nucleotide-binding motif A (P-loop)
F:585-586/Region: actin binding status predicted
F:685-1113/Domain: actin binding status predicted
F:685-913/Region: basic
F:920-968/Region: alanine/glycine/proline-rich
F:969-1016/Domain: SH3 homology <SH3>
F:1016-1113/Region: alanine/glycine/proline-rich
F:107-binding site: ATP (lys) #status predicted
Query Match 8.2%; Score 176.5; DB:1; Length 1113;
Best Local Similarity 21.9%; Pred. No. 0.0057;
Matches 89; Conservative 65; Mismatches 148; Indels 105; Gaps 20;
QY 71 NSLRFSFSLKQOTENVSAHTQALALREBELRSLEFREROK--EORRKYEAIMDR---- 125
Db 652 SGVEITLTKMDLEPKOYSKGTIKIFRAPELVNLEELREKVFYRNKLQRPFLFTLM 711
QY 126 -----VQSKLSLYKKTMSKRAYDOKCDAD--DAEQAFER--VSANGCOK-----QVE 171
Db 712 SYIYSQKGAADSMKSNKERRRLSTERYQSDYINRENPELKDLYKKNGEKIMETHAV 771
QY 172 KQONNAKQCKES-----ATEERVRONIEQLEARTEW--EOEHR-----T 211
Db 772 NKYDRSRQGRVLLLDSTAIYFIATE-----KKKDKDRKKRPYIYQKRLLAGIT 825
QY 212 TCEAQLQLEFDRLTILRNALWYHCHOLSMQCVKDELYEVRLTLEGCDVEGDING-FIQ 270
Db 826 SYELSKLS--DGFVYL-----RNNEHQIFE-----CRKTEFLGTILK 863
QY 271 SKSTGRPPAPVPYQNYDREVP-----LIGSPISQPCGYIKRSGILHSP 319
Db 864 AYKIG---TLRINTNNSIGVAIKASKQGGKGERITLEFKGIKPESEVFK-----GTK 913
QY 320 KTTBS--APAASTETLTPTPERNELVYASIEVQATQGNLSSAQD-----YAL 366

Db 914 VSTPSDGLPADTVNLTLP--PE--SLPVYSIPILYKPAAMAKNAPONSQSPASNVKPSAKAL 970
QY 367 YDYTAONSDELIDISAGDILAVILEGEGDMMTVERNQORGFPVPGSYLE 413
Db 971 YDFAESSMELSFKEGDILVLDOSSGDMMDAELKGRGVPSNYLQ 1017
RESULT 13
T23452
hypothetical protein K08E3.3a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T23452
R:McMurray, A.
Submitted to the EMBL Data Library, November 1996
A:Reference number: 219743
A:Accession: T23452
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-783 <WIL>
A:Cross-references: EMBL:281568; PIDN:CAB04591.1; GSPDB:GN00021; CESP:K08E3.3a
A:Experimental source: clone K08E3
C:Genetics:
A:Gene: CESP:K08E3.3a
A:Map position: 3
A:Introns: 43/3; 99/3; 133/3; 186/3; 225/2; 266/3; 320/2; 355/3; 424/1; 451/2; 477/1;
Query Match 8.0%; Score 173; DB:2; Length 783;
Best Local Similarity 18.7%; Pred. No. 0.006;
Matches 114; Conservative 81; Mismatches 194; Indels 222; Gaps 20;
QY 14 RDTATGTEVLLQRLDGRKCKDVEELLROPAQAEERYKELVOIARAGQTEM-- 71
Db 181 RFTYVODPSNALSTYQKIDFEFEKLGQFSKEKAAIEEYSTRKSLAKKVKAKKSPEDDE 240
QY 72 -----SLRTSFDSLKQOTENVSAHTQALALREBELRSLE-----EFREROKRKTYA 121
Db 241 ILKSVSYTSSFNFLQDQIATRIHOTSAEHIRGGVASTVASKTCOMRSRNAINDKT 300
QY 122 IMDRVQSKLSLYKKTMSKRAYDOKCDADAEQAFERVSANGHOKQVEKQONKAKOCK 181
Db 301 INDKLEDDQ-----INEMKSGKCYLKSFKDAENSQKFYVADKNLEISRL--LEKARALA 354
QY 182 ESATEAERYRONIEQLERARTEMQDEHNTTCEAQLQLEFDLTLIRNALWYHCHOLSMQ 241
Db 355 NARNEACELAKQDYSAVLR--KTNAEOK--RYHVELLPV--IFARLKAVDRECIADMRQVLOK 411
QY 242 CVK--DDEL--YEEVRLTLE-----GQVEGDINGFIQSKSTGRPPAPVPYQNYDREVT 293
Db 412 IVSFDSILADSYEECEKRIQOREVGKIDAEQDQVLKSVENATIEQPAFIEIDLGDPKNC 471
QY 294 PLIGSPISQPCG-----VIKRPSGLL----- 315
Db 472 DSRTNDSADGSGKLLKSSPSKNRIIRNFLGLKEKREADEKPEASNNQDLMTYDKSPAH 531
QY 316 -----HGSPK----- 320
Db 532 VRLSCLRSKIRNMEQLOALQGRGILRLQOAYTNPQHGHPACTEPLISYAKKLEKL 591
QY 321 ----- 320
Db 592 KMDIHLKEFYAMLEMSVEEGQERFGGRTPTTRMSGSSSTNOSSKTIEDVLSGAG 651
QY 321 -----TTP--SAPASITETLTPTPERNELVYASIEV--QATQGNLN 357
Db 652 NSSADSSKNILNQLFTTPKRLISSPRTS--KSSTPPLRRRAISSPKILRSSFSGAIR 710
QY 358 SS-----ADYRALYDYTAONSDELIDISAGDILAVILEGE--DGWTVYER--NG 402
Db 711 KSLSTPQSVKVEYAVTYALFEEFAKSSAETMSIEQGLL--LVLEHDHGDGWTRTKKRHN 769
QY 403 QRGFVPGSYLE 413

QY 401 NGORFVPGSTYLE 413
| : ||||| :
Db 770 NESGFVPTSTYLQ 782

Search completed: July 28, 2003, 09:16:44
Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2003, 09:15:23 ; Search time 15 Seconds
(without alignments)
1301.072 Million cell updates/sec

Title: US-09-068-377C-1
Perfect score: 2152
Sequence: 1 MNAQLQPRDAFWCRDFTAHT.....WTVERNGRGFPVPSYLEKL 415

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	737.5	34.3	333	1	PSP2_MOUSE
2	729	33.9	333	1	PSP2_MOUSE
3	346	16.1	670	1	YBY2_SCHPO
4	345.5	16.1	927	1	CC15_SCHPO
5	284.5	13.2	488	1	PAC2_MOUSE
6	283.5	13.2	486	1	PAC2_MOUSE
7	276.5	12.8	444	1	PAC1_HUMAN
8	276.5	12.8	486	1	PAC2_MOUSE
9	275	12.8	424	1	PAC3_MOUSE
10	268	12.5	424	1	PAC3_MOUSE
11	266.5	12.4	441	1	PAC1_MOUSE
12	266.5	12.4	448	1	PAC2_MOUSE
13	265.5	12.3	441	1	PAC1_MOUSE
14	259	12.0	477	1	PAC2_MOUSE
15	238.5	11.1	412	1	GAS7_MOUSE
16	236.5	11.0	421	1	GAS7_MOUSE
17	227	10.5	422	1	GAS7_MOUSE
18	206	9.6	1217	1	ITN1_MOUSE
19	202	9.4	1721	1	ITN1_MOUSE
20	198	9.2	1714	1	ITN1_MOUSE
21	194.5	9.0	1270	1	ITN1_MOUSE
22	191.5	8.9	416	1	EG13_ECHGR
23	191.5	8.9	1696	1	ITN2_MOUSE
24	188.5	8.8	1658	1	ITN2_MOUSE
25	177.5	8.2	426	1	EM13_ECHMU
26	176.5	8.2	1113	1	EM13_ECHMU
27	175.5	8.2	1113	1	EM13_ECHMU
28	171	7.9	1168	1	EM13_ECHMU
29	167	7.8	546	1	SRB8_MOUSE
30	166.5	7.7	633	1	SRB8_MOUSE
31	166	7.7	550	1	SRB8_MOUSE
32	165	7.7	526	1	SRB8_MOUSE
33	164.5	7.6	1014	1	NEBL_HUMAN

34	163.5	7.6	2477	1	SPCN_CHICK	P07751 gallus gall
35	162	7.5	527	1	NCF2_BOVIN	O77775 bos taurus
36	161	7.5	527	1	SPCN_HUMAN	O13813 homo sapien
37	160.5	7.5	642	1	YB65_SCHPO	O09746 schizosacch
38	160.5	7.5	777	1	YAS1_SCHPO	O09697 schizosacch
39	159.5	7.4	482	1	R167_YEAST	P39743 saccharomyc
40	159.5	7.4	1109	1	MYIE_HUMAN	O12965 homo sapien
41	159	7.4	2472	1	SPCN_MOUSE	P16086 rattus norv
42	158	7.3	563	1	SRB8_CHICK	O01406 gallus gall
43	158	7.3	1920	1	PCT2_MOUSE	P48725 mus musculu
44	157.5	7.3	1107	1	MYIE_MOUSE	O63356 rattus norv
45	151.5	7.0	2415	1	SPCA_DROME	P13395 drosophila

ALIGNMENTS

RESULT 1
PSP2_MOUSE STANDARD; PRT; 333 AA.
AC 099M15; 092189;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Proline-serine-threonine phosphatase-interacting protein 2 (Macrophage
DE actin-associated tyrosine-phosphorylated protein) (pp37).
GN PSTPIP2 OR MAIP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-9; 51-60; 75-80; 103-110;
RP 116-127; 172-175; 181-195; 200-209; 223-227; 263-270 AND 272-286.
RX MEDLINE=99023997; PubMed=9804836;
RY Yeung Y.-G., Soldera S., Stanley E.R.;
RT "A novel macrophage actin-associated protein (MAIP) is tyrosine
RT phosphorylated following colony stimulating factor-1 stimulation.";
RT J. Biol. Chem. 273:30638-30642(1998).
RN [2]
RP SEQUENCE FROM N.A.
RP TISSUE=Mammary gland;
RX MEDLINE=22388257; PubMed=12477932;
RA Strusberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schler G.D.,
RA Altschuld S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Locantore N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butlerfield V.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RL -i- FUNCTION: Blinds to F-actin. May be involved in regulation
CC of the actin cytoskeleton.
CC -i- SUBCELLULAR LOCATION: Cytosolic and membrane fractions.
CC -i- TISSUE SPECIFICITY: Expressed in macrophage-containing tissues,
CC including bone marrow, spleen, liver, kidney, intestine and brain.
CC -i- PTM: Phosphorylated on tyrosine.
CC -i- SIMILARITY: Contains 1 FCH domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; Y18101; CAAT7027.1; -
 DR EMBL; BC002123; AA02123.1; -
 DR MGI; 1335088; Pstpip2.
 DR GO; 0005856; C:cytoskeleton; IDA.
 DR GO; 0005829; C:cytosol; IDA.
 DR GO; 0005624; C:membrane fraction; IDA.
 DR GO; 0003779; F:actin binding activity; NAS.
 DR GO; 0007011; P:regulation of cytoskeleton; NAS.
 DR InterPro; IPR001060; Cdc15_Fes_CIP4.
 DR Pfam; PF00611; FCH; 1.
 DR SMART; SM00055; FCH; 1.
 DR PROSITE; PS50133; FCH; 1.
 DR Colled coil; Phosphorylation.
 FT INIT_MET 0
 FT DOMAIN 3 88 FCH.
 FT COLLED COIL (POTENTIAL).
 FT CONFLICT 65 162 E -> G (JIN REF. 1).
 FT SEQUENCE 297 297
 SQ SEQUENCE 333 AA: 38889 MW: A045F3FD2140FE6 CRC64;

Query Match 34.3%; Score 737.5; DB 1; Length 333;
 Best Local Similarity 40.1%; Pred. No. 1,1e-35;
 Matches 146; Conservative 67; Mismatches 112; Indels 39; Gaps 5;

QY 7 FRAAFKCRDFTAHGTEVLLQRLDGRKCKDVELLRQRAQAEERYGKELVQIARRAG- 65
 DB 5 FKGFNFWSTDLITSTGYSIIOHLNNGRKNCKEPEDFEKRAISIEKYGKDLNLNRRKPC 64
 QY 66 GQTEMSLRTPSESLKOOTENVSASHIOLALREELRSLEPERROCKEKKYEALMDR 125
 DB 65 GQSIINTLKRLEFFKQOVAVACHIQLAOTLEERKKEEFKEKQDLQKKTETMDA 124
 QY 126 VQSKSLLYKTKMESKRAYDQCDADAEAFERVSANGHOKQVEKSONKAKCKESAT 185
 DB 125 AHKQRMQFQKAMDAKKNYEOKRCDDEAQAIVRSANVANQROEKLTVKLAATKTAVE 184
 QY 186 EAEVNYKQNTIEQLERATEMEQEHRTCEAFQLOEFRLTLRLNALVHCHNLSMOCVKD 245
 DB 185 DSDAVALHMLNLEKVEEDQSEHAKCEVFEACERINFRNALHLNQLSQCCVAN 244
 QY 246 DELVEEVRLEFGDGVEDNGFISQSTGREPPAPYQVYQVREVTPLIGSPSIQSC 305
 DB 245 DEWEQYRKSLSTICKDIQYFNQKRTGOTPPAPIMYENFY-----SPQ----- 290
 QY 306 GVTKRFSGLLHSGPKTTPSAPAASTETLPTPER--NELVYASIEVQATGNNLSAODY 363
 DB 291 -----RNAAPERKTGPNPARRGLPLVPRKRPDPDPYSVE-----DY 328
 QY 364 RALY 367
 DB 329 SLTY 332

RESULT 2
 PSP2_HUMAN
 ID PSP2_HUMAN STANDARD; PRT; 333 AA.
 AC 09H939;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Proline-serine-threonine phosphatase-interacting protein 2.
 GN Pstpip2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NX NCBI_TaxID=9606;
 RN [1]
 SEQUENCE FROM N.A. (ISOFORM 2).

RC TISSUE-Testis;
 RX MEDLINE-22388257; Pubmed-12477932;
 RA Strausberg R.L., Reinhold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loguettano N.A., Peters G.J., Abramson R.D., Mulhany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Wooley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalios D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [2]
 RP SEQUENCE OF 112-333 FROM N.A. (ISOFORM 1).
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
 RA Nagatsuma M., Hosokiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Yamamoto J., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Ninomiya K., Iwananagi T.;
 RT "NEDO human cDNA sequencing project";
 RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Binds to F-actin. May be involved in regulation of the
 CC actin cytoskeleton (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytosolic and membrane fractions (By
 CC similarity).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q9H939-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9H939-2; Sequence=VSP_004070;
 CC Note=No experimental confirmation available;
 CC -1- PTM: Phosphorylated on tyrosine (By similarity).
 CC -1- SIMILARITY: Contains 1 FCH domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; AK023100; BAB14404.1; ALT_INIT.
 DR EMBL; BC035395; AAH35395.1; -
 DR Genew; HGNC:9581; Pstpip2.
 DR InterPro; IPR001060; Cdc15_Fes_CIP4.
 DR Pfam; PF00611; FCH; 1.
 DR SMART; SM00055; FCH; 1.
 DR PROSITE; PS50133; FCH; 1.
 DR Colled coil; Phosphorylation: Alternative splicing.
 FT DOMAIN 3 88 FCH.
 FT COLLED COIL (POTENTIAL).
 FT CONFLICT 65 165 MEQVRSLEMSIQDIEFVNQKRTGOTPPAPIMYENFY
 FT VARSPLIC 247 333 MSQKNVPAQKATGPNLARGPLIPKSSPDSPNSLYDY
 FT SLTY -> HPSCMRISTPPRRMSQOERLQGLTWQGEPS
 FT OFLKAHOMJPLTLMWTVCSTSKINETRAFGS (In
 FT isoform 2).
 FT /FTId=VSP_004070.
 FT SEQUENCE 333 AA: 38701 MW: 0A8C9FD4C419A2B8 CRC64;


```

Db 547 PASGTAPDDEPYTRDSSFFVDDTINTKATGNTSNRLSPAPPTDGGDTSIDNPTSDGQ 606
Qy 363 -----YRALXYTQNSDELDSAGDILAVILEGDDGW-----IYERNGOGFVPGVSGYLE 413
Db 607 RILGVSLALYDIDAIPETISFRKDTIAVLKLYEDGWEGFVGEDDHNRGQFSPNFR 666
Qy 414 KL 415
Db 667 EI 668

RESULT 4
CC15_SCHPO STANDARD: PRT: 927 AA.
AC 009822: 014365: 09050:
DT 01-FEB-1996 (Rel. 33, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cell division control protein 15.
GN CDC15 OR SPAC20G8.05C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=95360987; PubMed=7634333;
RA Fankhauser C., Reymond A., Cerutti L., Utzig S., Hofmann K.,
RA Simanis V.;
RT "The s. pombe cdc15 gene is a key element in the reorganization of F-
RL actin at mitosis."
RL Cell 82:435-444(1995).
RN [2]
RP REVISIONS TO N-TERMINUS.
RA Fankhauser C., Reymond A., Cerutti L., Utzig S., Hofmann K.,
RA Simanis V.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouris J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Goules S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
RA Jones K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Voickaert G., Aert R., Robben J., Grymoprez B.,
RA Weltjens J., Vaststreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Borzom C., Fuchs M., Filtz C., Holzer E., Moestl D., Hilbert H.,
RA Borzom K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Motier S.,
RA Gaibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Rhode G.,
RA Daza R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.B., Paulsen I., Potashkin J.,
RA Shpakovskii G.V., Ussery D., Barrett B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomycetes pombe."
RL Nature 415:871-880(2002).
RN [4]
RP SEQUENCE OF 286-425 FROM N.A.
RC STRAIN=968 h90;

```

```

RX MEDLINE=20223868; PubMed=10759889;
RA Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,
RA Hiraoka Y.;
RT "Large-scale screening of intracellular protein localization in living
RT fission yeast cells by the use of a GFP-fusion genomic DNA library."
RL Genes Cells 5:169-190(2000).
CC -!- FUNCTION: AFTER THE ONSET OF MITOSIS, FORMS A RING-LIKE STRUCTURE
CC WHICH CO-LOCALIZES WITH THE MEDIAL ACTIN RING. APPEARS TO MEDIATE
CC CYTOSKELETAL REARRANGEMENTS REQUIRED FOR CYTOKINESIS. ESSENTIAL
CC FOR VIABILITY.
CC -!- DEVELOPMENTAL STAGE: PEAKS IN EARLY MITOSIS BEFORE SEPARATION.
CC -!- DOMAIN: THE N-TERMINAL REGION IS IN A COILED-COIL STRUCTURE.
CC -!- PTM: PHOSPHORYLATED.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X86179; CA60115.1;
DR EMBL: Z95334; CAB08599.2;
DR EMBL: AB027810; BAA87114.1;
DR PIR: T38127; T38127.
DR HSP, P07751; ITUD.
DR GenedB_Spombe: SPAC20G8.05C;
DR InterPro: IPR001060; Cdc15_Fes_CIP4.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00018; SH3; 1.
DR Pfam: PF00611; FCH; 1.
DR PRINTS: PR00452; SH3DOMAIN.
DR PRODOM: PD000066; SH3; 1.
DR SMART: SM00055; FCH; 1.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS50133; FCH; 1.
DR PROSITE: PS50002; SH3; 1.
DR Mitosis: Cytoskeleton; SH3 domain; Phosphorylation; Coiled coil.
FT DOMAIN 20 101 FCH.
FT FT 108 207 COILED COIL (POTENTIAL).
FT DOMAIN 866 927 SH3.
SQ SEQUENCE 927 AA; 102119 MW; FDCE7ED0AA3D247D CRC64;

Query Match 16.18; Score 345.5; DB 1; Length 927;
Best Local Similarity 25.98; Pred. No. 1e-12;
Matches 96; Conservative 74; Mismatches 169; Indels 31; Gaps 8;

Qy 3 AOLQFRAPFCRDPHTAHTGVEVLRLDGRKKCKDVEELLRQRAOAEERYGKELVQIAR 62
Db 18 SVKFRNFNGSE---DAGDALSRRKSSLSVLESIDETAKRASIERYASKLOELAA 74
Qy 63 KAGQGTENSLRTSFDLSKQOTENVGSANIQLALALNEELRS--LEEFRRQKQRRKYEA 121
Db 75 SSADIPVGVSTLNNILISMRTGSMKAHBEVSGQINTLRKRIREVTDTEQKVAAN 134
Qy 122 IMDRVQSKSLYKTKYTESKAYDQCKRADDADQAFERYSANGHOKOVESQNNKKQCK 181
Db 135 AIELVYOKTKALEIDISEKDAVEYSCNKNLSYRQTKWMTG---RELDKVMKLRQAA 190
Qy 182 ESATEARVYRONTEOLERARTMEQEHRTTCAPFOLEDRITLIRNLMLVNCNOLSMQ 241
Db 191 LAYKKDAEYREINELLITRTREIDRWTEVCAFOHIEYRLLEFLKTNMAYANLISIA 250
Qy 242 CAVDDLELYEERLTLECCDEVCIDNGFTQSKSTGREGPAVPYQNTYDRE-----VTP 295
Db 251 CVMDESCERIRLTLENTINDEDITQIMQWEGCTIIP-PLPFPNDYFKENGNGINTYD 309
Qy 296 IGSFSDIPSCGVTKRPSGLLHSGPKTTPSAAPASTETLT-----PT-----PERN 340
Db 310 ISKAPSYPSYSSSRPASASIASLASSPTRSAFPRKTS-ETVSSEVVSPTSPLHSPVAP 368

```

OY 341 ELVASTIEVQ 350
Db 369 EOVEQVTEVE 378

RESULT 5

PAC2_RAT
ID PAC2_RAT STANDARD: PRT: 488 AA.
AC 09QY17; 09QY18; 09QY19; 09QY20;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein kinase C and casein kinase substrate in neurons 2 protein
DE (synaptic dynamin-associated protein II) (Synadin 2) (Synadin-II)
DE (SOP11).
GN PACSIN2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3 AND 4).
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=20171438; PubMed=10704453;
RA Qualmann B., Kelly R.B.;
RT "Synadin isoforms participate in receptor-mediated endocytosis and
RT actin organization."
RT J. Cell Biol. 148:1047-1062(2000).
RN [2]
RP FUNCTION.
RX PubMed=11082044;
RA "Modregger J., Ritter B., Witter B., Paulsson M., Plomann M.;
RT "All three PACSIN isoforms bind to endocytic proteins and inhibit
RT endocytosis."
RL J. Cell Sci. 113:4511-4521(2000).
CC - FUNCTION: May play a role in vesicle formation and transport.
CC - SUBUNIT: Homo- and hetero-aggregates with other PACSINs. Binds
CC dynamin 1, synaptojanin, synapsin 1 and the neural Wiskott-Aldrich
CC syndrome protein (N-WASP).
CC - SUBCELLULAR LOCATION: Cytoplasmic. Vesicle-like cytoplasmic
CC distribution.
CC - ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=4;
CC Name=1; Synonyms=AA;
CC Name=2; Synonyms=AB;
CC Name=3; Synonyms=BA;
CC Name=4; Synonyms=BB;
CC Name=4; Synonyms=BB;
CC IsoId=09QY17-4; Sequence=VSP_004518; VSP_004519;
CC - TISSUE SPECIFICITY: Ubiquitously expressed. Isoforms 1 and 3 are
CC expressed in P12 cell line and heart, whereas isoforms 2 and 4
CC are seen in most tissues examined with higher levels of expression
CC in muscle, testis and brain.
CC - PTM: Phosphorylated by casein kinase 2 (CK2) and protein kinase C
CC (PKC) (by similarity).
CC - SIMILARITY: BELONGS TO THE PACSIN FAMILY.
CC - SIMILARITY: Contains 1 FCH domain.
CC - SIMILARITY: Contains 1 SH3 domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF139492; AAF22211.1;
CC DR EMBL: AF139493; AAF22212.1;
CC DR EMBL: AF139494; AAF22213.1;
CC

DR EMBL: AF139495; AAF22214.1;
DR HSSP: P29355; ISEM.
DR InterPro: IPR001060; Gccl5_Fes_C1P4.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00611; FCH; 1.
DR Pfam: PF00018; SH3; 1.
DR PRINTS: PR00452; SH3DOMAIN.
DR ProDom: PD000066; SH3; 1.
DR SMART: SM00055; FCH; 1.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS50133; FCH; 1.
DR PROSITE: PS50002; SH3; 1.
KW Endocytosis; SH3 domain; Coiled coil; Phosphorylation;
KW Alternative splicing;
FT DOMAIN 11 75 FCH.
FT DOMAIN 428 488 SH3.
FT DOMAIN 184 239 COILED COIL (POTENTIAL).
FT VARSPPLIC 302 303 Missing (in isoform 4).
FT VARSPPLIC 346 386 Missing (in isoform 2 and isoform 4).
FT SEQUENCE 488 AA; 55978 MW; B2975012EF0FDP56 CRC64;

Query Match 13.2%; Score 284.5; DB 1; Length 488;
Best Local Similarity 21.8%; Pred. No. 1.6e-09;
Matches 107; Conservative 87; Mismatches 194; Indels 103; Gaps 15;

OY 9 DAFRCRPTMTG-YEVLLRLDGRKMKDYELLROAQAERYKELVQIARRAGQ 67
Db 16 DFW-----EYGNKRTVKRIDDHRLCGDLNCLHERARIEKAYAOQLTWMARWRL 69
OY 68 TE-----MNSLRSPSLQOQTENVSANHTALALREELRSLEEFPERKE----- 114
Db 70 VEKGQYQIVKAKMAVNSEAEVSELSLEVASLNE--DFEKTINMCKEAFHKOMGC 127
OY 115 -QRRKYEAINDRVQSKSLTYKTMESKAYDOKCRADDA--DQAFRVANGHOKOV 170
Db 128 FKRETEADGPRKAKPAKKIKLEVDAAKKAHNTACKKEKLAVNSRANKAPSLNPEL 187
OY 171 EKSQKAKCKESATEAEVRVQNTQLEARRTEVDQEHRTTCGAFQLOEPRRLTLRNA 230
Db 188 KKLQPKIKCKQDVKTQDKYKAKLELDQTPQYMNENQVFEQCOQFEERLRFREV 247
OY 221 LWHGNOLSMQCVKQDE-LYEEVRLLECGDVEGNGTQSKSGREPPAVPQNYVD 289
Db 248 LLEVCKHLDSLNVASYKGYRLEQSIKADAVEDIRNF-----RANHGFAMAMNQFED 303
OY 290 RE-----VTPPLIGSPSTQPSGVYIKRPSGLHGSPTTPSAPAS 329
Db 304 EEMSAIDLNRITLSRRKKKAADGV-TLTGINQCGDQSGQKPPSNL-----SVPSNPAGS 356
OY 330 T-----ETLPTPERNELVYASIE-----VQATQGL 356
Db 357 TOLQSSYNPFDEDDTSSVSEKEDIKAMVSYKQNYVPADWSDSNPNFSSDANG 416
OY 357 NSSAGD-----YRALVYTAQNSDELDISAGDILAVLLEBDGWYTER--NGOR 404
Db 417 DSNPFDEDTTSTEVKRALVDYEGQEHDELFSKAGDELTKLEDEDDGCKGRDGOV 476
OY 405 GVPQSVYLEKL 415
Db 477 GLYPANYVEAI 487
RESULT 6
PAC2_MOUSE
ID PAC2_MOUSE STANDARD: PRT: 486 AA.
AC 09WYB8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Protein kinase C and casein kinase substrate in neurons protein 2.
GN PACSIN2.

OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6 x DBA;
 RX MEDLINE=99358785; PubMed=10431838;
 RA Rittler B., Modregger J., Paulsson M., Plomann M.;
 RT "PACSIN 2, a novel member of the PACSIN family of cytoplasmic adapter
 proteins.";
 RL FEBS Lett. 454:356-362(1999).
 RP SEQUENCE FROM N.A.
 RC TISSUE=Mammary gland;
 RX MEDLINE=23388257; PubMed=12477932;
 RA Strusberg R.L., Feligold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stacione M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loguailano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schultz J., Myers R.M.,
 RA Bottingfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
 RA Scheraga A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP FUNCTION.
 RX PubMed=11082044;
 RA Modregger J., Rittler B., Witter B., Paulsson M., Plomann M.;
 RT "All three PACSIN isoforms bind to endocytic proteins and inhibit
 endocytosis.";
 RL Cell Sci. 113:4511-4521(2000).
 CC -1- FUNCTION: May play a role in vesicle formation and transport.
 CC -1- SUBUNIT: Homo- and hetero-aggregates with other PACSINs. Binds
 CC dynamin 1, synaptobrevin, synapsin I and the neural Wiscott-Aldrich
 CC syndrome protein (N-WASP) (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic. Vesicle-like cytoplasmic
 CC distribution (By similarity).
 CC -1- TISSUE SPECIFICITY: Ubiquitously expressed, with highest levels in
 CC brain, heart, skeletal muscle, and ovaries.
 CC -1- PTM: Phosphorylated by casein kinase 2 (CK2) and protein kinase C
 CC (PKC) (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE PACSIN FAMILY.
 CC -1- SIMILARITY: Contains 1 SH3 domain.
 CC -1- SIMILARITY: Contains 1 SH3 domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF128535; AADA1780.1; -;
 DR EMBL: BC023502; AAH23502.1; -;
 DR HSSP: P29355; ISEM.
 DR MGD: MGI:1345153; Pacsin2.
 DR GO: GO:0005829; C:cytosol; IDA.
 DR GO: GO:0008092; F:cytoskeletal protein binding activity; IDA.
 DR GO: GO:0045806; P:negative regulation of endocytosis; IDA.
 DR InterPro: IPR001060; Cdc15_Fes_C1P4.

DR InterPro: IPR001452; SH3.
 DR Pfam: PF000611; FCH; 1.
 DR Pfam: PF000018; SH3; 1.
 DR PRINTS: PR00452; SH3DOMAIN.
 DR PRODOM: PD000066; SH3; 1.
 DR SMART: SM00055; FCH; 1.
 DR SMART: SM00326; SH3; 1.
 DR PROSITE: PS00133; FCH; 1.
 DR PROSITE: PS00002; SH3; 1.
 DR Endocytosis; SH3 domain; coiled coil; phosphorylation.
 FT DOMAIN 11 75 FCH.
 FT DOMAIN 426 486 SH3.
 FT DOMAIN 184 239 COILED COIL (POTENTIAL).
 SQ SEQUENCE 486 AA; 55833 MW; 66C17ECC0767E0E7 CRC64;
 Query Match 13.2%; Score 283.5; DB 1; Length 486;
 Best Local Similarity 21.5%; Pred. No. 1.8e-09;
 Matches 106; Conservative 86; Mismatches 193; Indels 107; Gaps 14;
 OY 9 DAFRCRPTAHTG-YEVLLORLDRGRMKCKVDELLRQRAQAEERYGKELVQIARKAGQ 67
 DB 16 DSEF-----EVGNYKRTVKRIDGCHRLCGDLNMLERARIKAYAOULTENARRRQL 69
 OY 68 TE-----MNSLTSPDSLKQOTENVGSAHIALALRELSLEFRERORE----- 114
 DB 70 VEKPGQGTVEKAWIAVMSAEARVSEHLLEVKASLME--DFEIKVMQKEAFHKQMMG 127
 OY 115 --QKKKEALMDRQKSKLSYKKTMTSKAYIDCKDADDA--EQAFERYANGHOKOV 170
 DB 128 FKREKEADGFRKQKQWAKKLEVEAKRAHHTACKEEKLAISREANSKADSLNPEOL 187
 OY 171 EKSNKKKCKESKSTAEERYRONIEQLERARTEMEQEHRTQCEAFOLQEDRLTILRNA 230
 DB 188 KKLDQDKIEKQVDLTKDKYKESKLKELDQTPPYIMNMQVPECCQFEKRLRFREV 247
 OY 231 LNVHCNLSMQCKVD-DELYEVRLLTEGCDVEGDINGFIQSKSTG----- 275
 DB 248 LLEVQKHLDSLNVASVYKTYIRELEQSIKADAVEDELFWFRANRPGMAMWMPQEEWSAD 307
 OY 276 -----REPPAPVYQNYNDREYTPLGSPSTQSGCVIKRRSGLLHSGPKTTPSPAPA 328
 DB 308 LNRLSRREKKKAV-----DGVTLTGINQGDQSGQNK-----PGSNLSVSPAPQ 353
 OY 329 ST-----ETLTPPERNELVVASIE-----VOATQGN 355
 DB 354 STQAGSYNPFEDDDTGGSSISEKEDKANNVSEYKQTYPPIDWSDDESNPFSSIDAN 413
 OY 356 LNSSAQD-----YRALYDTAONSDELDISAGDILLAVIEGEGWTVVER--NGQ 403
 DB 414 GDSNPFEDDTTSGTEVRVRLALDYEGQEHDELSPKADDELTKIEDDEQGMCKGRUDSGQ 473
 OY 404 RGFVPGSYLEKL 415
 DB 474 VGLYPAWTVETI 485
 RESULT 7
 PAC1_HUMAN STANDARD; PRT; 444 AA.
 AC Q9BYL1; Q9P268;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Protein kinase C and casein kinase substrate in neurons protein 1.
 CN PACSIN1 OR KIAA1379.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21100457; PubMed=11179684;
 RA Sunoy L., Pluvinet R., Andreu N., Estivill X., Escarceller M.;

```

RT "PACSIN3 is a novel SH3 domain cytoplasmic adapter protein of the
RT pascin-syndaplin-PAP32 gene family.";
RL Gene 262:199-205(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20181126; PubMed=10718198;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirosewa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVI.
RT The complete sequences of 150 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:65-73(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Amygdala;
RX Oltjenweider B., Obermaier B., Mewes H.-W., Weil B., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP FUNCTION;
RX PubMed=11082044;
RA Motteger J., Rilter B., Witter B., Paulsson M., Plomann M.;
RT "All three PACSIN isoforms bind to endocytic proteins and inhibit
RT endocytosis.";
RL J. Cell Sci. 113:4511-4521(2000).
CC -i- FUNCTION: May play a role in vesicle formation and transport.
CC -i- SUBUNIT: Homo- and hetero-aggregates with other PACSINs. Blinds
CC dynamin I, synaptojanin, synapsin I and the neural Wiskott-Aldrich
CC syndrome protein (N-WASP) (By similarity).
CC -i- SUBCELLULAR LOCATION: Cytoplasmic. Colocalizes with dynamin I at
CC vesicular structures in the cell body and neurites (By
CC similarity).
CC -i- TISSUE SPECIFICITY: Highly expressed in brain and in a lesser
CC extent in heart, pancreas and liver.
CC -i- PTM: Phosphorylated by casein kinase 2 (CK2) and protein kinase C
CC (PKC) (Probable).
CC -i- SIMILARITY: BELONGS TO THE PACSIN FAMILY.
CC -i- SIMILARITY: Contains 1 FCH domain.
CC -i- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF242529; AAK92206.1; -
DR EMBL: AB037800; BAA92617.2; ALT_INIT.
DR EMBL: AL834211; CAD38895.1; -
DR HSSP; 060631; 1GB0.
DR Genew: HGNC:8570; PACSIN1.
DR MIM; 606512; -
DR GO; GO:0005737; C:cytoplasm; NAS.
DR GO; GO:0005515; F:protein binding activity; NAS.
DR GO; GO:0004672; F:protein kinase activity; NAS.
DR GO; GO:0006897; P:endocytosis; NAS.
DR InterPro: IPR001060; GcG15_Fes_C1P4.
DR InterPro: IPR001452; SH3.
DR Pfam; PF00611; FCH; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR PRODOM; PD000066; SH3; 1.
DR SMART; SM00055; FCH; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PSS0133; FCH; 1.
DR PROSITE; PSS0002; SH3; 1.
DR Endocytosis; SH3 domain; Coiled coil; Phosphorylation.
FT DOMAIN 13 76 FCH.
FT DOMAIN 147 168 COILED COIL (POTENTIAL).
FT DOMAIN 184 220 COILED COIL (POTENTIAL).
SQ SEQUENCE 444 AA; 50965 MW; 6AFAF801873770975 CAC64;

```

Query Match Similarity 12.8% Score 276.5; DB 1; Length 444;
Best Local Similarity 22.1%; Pred. No. 4e-09;
Matches 103; Conservative 83; Mismatches 197; Indels 83; Gaps 16;

```
DY      1  MAOLFRDAFWCRRDFTAHGTG-YEVLLQCLLGRKMKCKPVEEHLRQRAQAERNGKEIVOI 60  
DB      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
DY      2  LAPETTDSFV-----EVGNKKRVKRKHIDSHRCNDLMNCVQGRAIIEKAYGGQLTDW 63  
DB      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
DY      61  ARK-----AGCGTEMNSLRFSFDLSLKOOTENVSAAHILQALALAREELRSLEEFEROKE 114  
DB      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
DY      64  AKRMROLILEKGPO--YGSLERAMGAIMTEADKVESLHOEVKNMLINE--DEKYVMOKD 119  
DB      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
DY      115 -----QRKKYEALIMRVQCSKLSTLYKKTMEKSRAVDQCRADDAA--EGAFERY 161  
DB      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
DY      120 AYHKIIMGGEFKETTKAEDEGFRAQCPMAKKMIELPAARAAVHALACKEEKLMTMRMSNT 179  
DB      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
DY      162 SANGHQKQVKERSQNNAKCKESATFAEVRVQNIQEOLRRARTFMEOEHRRTCEAFLOEEF 221  
DB      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
DY      180 EGSVPPEQQKIKQDKVDCKQDVQCTQEKREYEVLEDVAKITPPQYNNEHQVFEGCCQPEE 239  
DB      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
DY      222 DRLTLIR-----NALMVHCNQLSMOCVKDELLEYEVRLTLEGCDVEGDING 267  
DB      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
DY      240 KRLVLAKFEVLLDIKRHNLIAENSYSIH-----YRELEQAIRADAQOEDLRW 286  
DB      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
DY      268 FIQSKSTGREPPAPRPYQNYVDRETYPLIGSPISLPSCGVIKRFSGL-LHSKPTTPEAP 326  
DB      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
DY      287 FRSTISGPQM--PMNPQPEEMEPMDLPHTTRKKEQP---KKAQVALTNATGAVESTS 339  
DB      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
DY      337 AASTETLTPPERNELVAASIEVQATQCN-----LNSSADY-----RALYDPTA 371  
DB      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
DY      340 QAGDRGSVSYDRQG-PYATEMSDESNGNPFPGSFTNGCANPFEDSDSGAVNRALYDVG 398  
DB      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
DY      372 QNSEFLDISAGDLILAVILEGEDGMVTVER--NGRGFPVGSYLEKL 415  
DB      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
DY      399 QEDELSEFRAGDELTKEEDEQGCGRGLRDSGGLTYPANYVEAI 444  
DB      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 8

PAC2_HUMAN

ID PAC2_HUMAN STANDARD; PRT: 486 AA.

AC Q9UNF0; O95921; Q96HV9; Q9HD03; Q9NP01; Q9Y4V2;

DT 28-FEB-2003 (Rel. 41, Created)

DI 28-FEB-2003 (Rel. 41, Last sequence update)

DI 15-SEP-2003 (Rel. 42, Last annotation update)

DE Protein kinase C and casein kinase substrate in neurons protein 2.
PACSIN.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_Taxid=9606;

RX [1]

RP SEQUENCE FROM N.A. (ISOFORM 1).

RC TISSUE=Brain, and Retina;

RA MEDLINE=99358785; Pubmed=10431838; PubMed=10431838; Blum H.,
LA Ritter B., Modregger J., Paulsson M., Plomann M.:
"PACSIN 2, a novel member of the PACSIN family of cytoplasmic adapter
RT proteins".
RN FEBS Lett. 454:356-362(1999).
[2]
RN SEQUENCE FROM N.A. (ISOFORM 2).

RC TISSUE=Testis;

RA MEDLINE=21154917; PubMed=11230166;
LA Wieman S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA Ansoerge W., Boecker M., Bloeker H., Bauersachs S., Blum H.,
LA Lauder J., Duesterhoeft A., Beyer A., Koehler K., Strack N.,
RA Mewes H.-W., Ottenwelder B., Obermayer B., Tampe J., Heubner D.,
LA Mamuth R., Korn B., Klein M., Posnetka A.;
RT "Towards a catalog of human genes and proteins: sequencing and
RN analysis of 500 novel complete protein coding human cDNAs".
LA Genome Res. 11:422-435(2001).
[3]
RN SEQUENCE FROM N.A.

FT CONFLICT 378 380 DDT -> EDI (IN REF. 1).
 SQ SEQUENCE 486 AA; 55738 MW; 821DBEF65DAD1AA8 CRC64;
 Query Match 12.8%; Score 276.5; DB 1; Length 486;
 Best Local Similarity 21.4%; Pred. No. 4.4e-09;
 Matches 103; Conservative 88; Mismatch 205; Indels 85; Gaps 14;
 OY 9 DAFMCRDPTAHNG-V-EVLLORLDRKMKCKVEELLROAQAEEYRGKELYIARKAGQ 67
 DB 16 DSFW-----EYGNKRTVKRIDGHRICSDMNCMLHEARLEKAYAOQLTEWARKROL 69
 OY 68 TE---MNSLRTSPDSLKQOTENWNSAHIQIALRLRELSLEFFERKE----- 114
 DB 70 VEKGPQYGVTEKAWMAFSEAEVSEELHLEVYKASLMND--DEKIKNNQKEAFHNMOMG 127
 OY 115 --QKKRYAIDRVOKSKLSLYKTKMESKAYADKCRDADA--QDAEYRSANGHOKOV 170
 DB 128 FKETKEADGGRKAKPWAKKLKEVEAKKAHNAKKEKLAISREANSKADPSINPEOL 187
 OY 171 EKSONKAKCKESATEAEYRYRONTEQERARTEMEQEHRTCEAFQOEPRILTLRMA 230
 DB 188 KKLDDKTEKQKODVLTKEKEKSLKEDDQTPQIMENNEQYFEOCCQFEERLRFRREV 247
 OY 231 LMVHCNOLSMQCKD-DELYEEVRLTLECGVEGDINGETIOSKSTGREPPAPVQNYD 289
 DB 248 LLEVQKHDLNSVAGYKAIYHDLQSIKRAADAVEDLRFMRANHGPMAMNPFQFEWSAD 307
 OY 290 -----REYPLIGSPISQSCYIKRPSGLIHSPPKT---PSAPASTETLT----- 334
 DB 308 LNRFLRSREKRKAFTDGVTL---TGINQGDQSLPKSPSTLWVPSNPASQSSQSNPF 364
 OY 335 -----PTPRNELVYASIE-----VOATQGLNSAAD----- 362
 DB 365 EDEDGTGTVSEKDTKAKNNSYKTKSTPTDMSDDSNFSSSTDANGSDNPFDDDAT 424
 OY 363 -----YRALDYTAQNSDELIDISAGILAVILEGDMWTVR--NGORGFPVPSYLEK 414
 DB 425 SGTREVRVALYDEQEHDELSEFKAGDELTKMEDQGWCKGRILNGCVGLYPANYVEA 484
 OY 415 L 415
 DB 485 I 485
 RESULT 9
 PAC3_MOUSE
 ID PAC3_MOUSE STANDARD: PRT: 424 AA.
 AC 09JUB8; Q9EOP9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Protein kinase C and casein kinase II substrate protein 3.
 GN PACSIN3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND FUNCTION.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=20534871; PubMed=11082044;
 RA Modregger J., Rilter B., Witter B., Paulsson M., Plomann M.;
 RT "All three PACSIN isoforms bind to endocytic proteins and inhibit
 RT endocytosis";
 RL J. Cell Sci. 113:4511-4521(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21100457; PubMed=11179684;
 RA Sunoy L., Pluviet R., Andreu N., Estivill X., Escarceller M.;
 RT "PACSIN 3 is a novel SH3 domain cytoplasmic adapter protein of the
 RT pacsin-syndapin-FAP52 gene family";
 RL Gene 262:199-205(2001).
 RN [3]

RP SEQUENCE FROM N.A.
 RC TISSUE-Neuron;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang Y., Hsieh F.,
 RA Diatchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mulhaly S.J.,
 RA Bosak S.A., Mcwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Heiton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalius D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- FUNCTION: May play a role in vesicle formation and transport.
 CC -1- SUBUNIT: Homo- and hetero-aggregates with other PACSINs. Binds
 CC dynamin 1, synaptojanin, synapsin 1 and the neural Wiskott-Aldrich
 CC syndrome protein (N-WASP). Also interacts with MDC9 and MDC15 (By
 CC similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic. Detected at the inner aspect of
 CC the plasma membrane in cells or myotubes (By similarity).
 CC -1- TISSUE SPECIFICITY: Highly expressed in skeletal muscle, heart and
 CC lung.
 CC -1- PTM: Phosphorylated by casein kinase 2 (CK2) and protein kinase C
 CC (PKC) (Probable).
 CC -1- SIMILARITY: BELONGS TO THE PACSIN FAMILY.
 CC -1- SIMILARITY: Contains 1 FCH domain.
 CC -1- SIMILARITY: Contains 1 SH3 domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-stb.ch).
 CC
 CC EMBL; AF149824; AAC31022.1;
 CC EMBL; AF242531; AAK29208.1;
 CC EMBL; BC003884; AAK03884.1;
 CC HSSP; P29355; ISEM.
 CC MGD; MGI:1891410; Pacsin3.
 CC GO; GO:0005737; C:cytoplasm; IDA.
 CC GO; GO:0008092; F:cytoskeletal protein binding activity; IDA.
 CC GO; GO:0045806; F:negative regulation of endocytosis; IDA.
 CC InterPro: IPR001060; Cdc15_Fes_CIP4.
 CC InterPro: IPR001452; SH3.
 CC Pfam; PR00611; FCH; 1.
 CC Pfam; PR00018; SH3; 1.
 CC PRINTS; PR00452; SH3DOMAIN.
 CC PRODOM; PD000066; SH3; 1.
 CC SMART; SM00055; FCH; 1.
 CC SMART; SM00326; SH3; 1.
 CC PROSITE; PS50133; FCH; 1.
 CC PROSITE; PS50002; SH3; 1.
 CC Endocytosis; SH3 domain; Coiled coil; Phosphorylation.
 CC DOMAIN 10 73 FCH.
 CC DOMAIN 363 424 SH3.
 CC DOMAIN 174 217 COILED COIL (POTENTIAL).
 CC CONFLICT 360 360 R -> G (IN REF. 1).
 CC SEQUENCE 424 AA; 48585 MW; 00475BC0321485B7 CRC64;
 Query Match 12.8%; Score 275; DB 1; Length 424;
 Best Local Similarity 22.1%; Pred. No. 4.6e-09;

Matches 97; Conservative 79; Mismatches 196; Indels 66; Gaps 12;

```

QY 10 AFWCDFAFHNGEYVLLDRLDGRKMCVDELLRQRAQAEERYKELVQARKAGGOTE 69
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 15 SFW-----EAGNYRTVQVREDEGHRLCDLVSCFERARIEKAAVQAQADWARKRGAVE 69
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 70 ----NSLRTSPDSLKQOTENWGSAHIOLALALREELSLFEFRERQEKQKYPAMD- 124
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 70 KGPQVGTLEKAMHAFETAERLSEHLT-----VRKKLNGPDSERVRTRMQRGAFHHPVLGG 125
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 125 -----RVQSKLSLYKKTKMESKKAIVDQKCRADDAE--QAEFRVSANGHQV 170
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 126 FRESRAADGGRKAKPWLKRLKEVEASKSHYARKMEKTAQRTESHAKADSSMSQDL 185
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 171 EKSQKAKCKESAEAEFRVYRONTEQERARTEWEOHRTTCFAFOJCEQDRLLTLNNA 230
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 186 RKLDERVGRCTKEAKKMTQYEQTLAEINRTTPRMEDEQAFESCQAERQRLFFEDV 245
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 231 LMVHGNOLSMOCV-KDELYEEVRLTEGCDVEGDIINGFIOSKSTGRPPAPVQNYD 289
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 246 LTLTHQHLDSLSSDKFELHRLDQSIASADEEDLRM-----RSTHGPAMMNPQ 298
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 290 REVPPLI-----GSPSTPSGCVTKRSGLLHGSEKTPPSAPASTETLPTP 337
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 299 FEESLDTQRAISRKEKGRSPDEVTLNIVPTRDGT--APPPOSPPSGSGODEWSD 356
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 338 ERNELVYASIEVQATQGNLNSADYRALYDTAQNSELDISAG-DILAVILEGEDEGM 396
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 357 ESPKRVATGVRY-----RALTYAGQEADELSPRAGEEELKMSDEDEQWC 402
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 397 TVE-RNGRGFPVPSYLE 413
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 403 QGOLQSGRIGLYPANYVE 420
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

```

RESULT 10

```

PAC3_HUMAN STANDARD: PRT: 424 AA.
ID PAC3_HUMAN
AC Q9UKS6; Q9H331; Q9NMV9;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Protein kinase C and casein kinase substrate in neurons protein 3 (SH3
DE domain-containing protein 6511) (Endophilin I).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Breast carcinoma;
RX MEDLINE=20002705; PubMed=10531379;
RA Howard L., Nelson K.K., Maciewicz R.A., Blobel C.P.;
RT "Interaction of the metalloproteinase disintegrins MDC9 and MDC15 with
RT two SH3 domain-containing proteins, endophilin I and SH3PXL.";
RL J. Biol. Chem. 274:31693-31699(1999).
RN [2]
RP SEQUENCE FROM N.A. AND FUNCTION.
RX PubMed=11082044;
RA Mooregger J., Rutter B., Witter B., Paulsson M., Plomann M.;
RT "All three PACSIN isoforms bind to endocytic proteins and inhibit
RT endocytosis.";
RL J. Cell Sci. 113:4511-4521(2000).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=211100457; PubMed=11179684;
RA Smuly L., Pluvinet R., Andreu N., Estivill X., Escarceller M.;
RT "PACSIN 3 is a novel SH3 domain cytoplasmic adapter protein of the
RT pectin-synadapin-FAP52 gene family.";
RL Gene 262:199-205(2001).
RN [4]
RP SEQUENCE FROM N.A.

```

```

RC TISSUE=Skin, and Uterus;
RX MEDLINE=22388257; PubMed=12477932;
RA Klausner R.D., Feinberg E.A., Grouse L.H., Derge J.G.,
RA Altschul S.F., Zeeberg B., Buelow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Ditchenko L., Marzina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schreier T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Cantinot P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huylk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Bakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [5]
RP SEQUENCE OF 1-405 FROM N.A.
RA Tanigami A., Fujiwara T., Ono T., Yamada K., Fujii Y., Ozaki K.,
RA Hirao M., Ohmori Y., Ota T., Suzuki Y., Obayashi M., Nishi T.,
RA Shibahara T., Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: May play a role in vesicle formation and transport.
CC -1- SUBUNIT: Homo- and hetero-aggregates with other PACSINs. Binds
CC dynamin 1, synaptophysin, synapsin 1 and the neural Wiscott-Aldrich
CC syndrome protein (N-WASP). Also interacts with MDC9 and MDC15 (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic. Detected at the inner aspect of
CC the plasma membrane in cells or myotubes.
CC -1- TISSUE SPECIFICITY: Expressed in heart and skeletal muscle,
CC lung, kidney, adrenal gland and placenta.
CC -1- PTM: Phosphorylated by casein kinase 2 (CK2) and protein kinase C
CC (PKC) (Probable).
CC -1- SIMILARITY: BELONGS TO THE PACSIN FAMILY.
CC -1- SIMILARITY: Contains 1 FCH domain.
CC -1- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF130979; AAF04472.1; -
DR EMBL: AF149825; AAG31023.1; -
DR EMBL: AF242530; AAK29207.1; -
DR EMBL: BC007914; AAH07914.1; -
DR EMBL: BC011889; AAH11889.1; -
DR EMBL: AK000577; BA091267.1; -
DR HSSP: P29353; ISEM.
DR Genew: HGNC:8572; PACSIN3.
DR MIM: 606513; -
DR InterPro: IPR001060; Cdc15_Fes_CIP4.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00611; FCH; 1.
DR PRINTS: PR00452; SH3DOMAIN.
DR PRODOM: PD000066; SH3; 1.
DR SMART: SM00055; FCH; 1.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PSS0135; FCH; 1.
DR PROSITE: PSS0002; SH3; 1.
DR Endocytosis: SH3 domain; Collated coll; Phosphorylation.
FT DOMAIN 10 73 FCH.

```

FT DOMAIN 363 424 SH3.
FT DOMAIN 167 217 COILED COIL (POTENTIAL).
FT CONFLICT 61 A -> V (IN REF. 1).
SQ SEQUENCE 424 AA: 48486 MW: 6DBD940AE6D1F352 CRC64:
Query Match 12.5%; Score 268; DB 1; Length 424;
Best Local Similarity 22.2%; Pred. No. 1,2e-08;
Matches 96; Conservative 76; Mismatches 205; Indels 56; Gaps 12;
QY 10 AFMCFDFAHTGYEVLLQRLDGRKMKVVEELLNORAOAEERYGELYOIRKAGGQIE 69
DB 15 SFW-----EAGNYRTVOVEDGHRICGDLVSCFERARIEKAYAOQLDARKMGTYE 69
QY 70 ----NNSRTSPDLSKQOTENGVSAHIQIALALREELSLSEFFEROKKRYEAIND- 124
DB 70 KPGYGTLEKAMHAFETAEERLSALHLE----VRELGQDSEDERVAMQRCAPHRVYLG 125
QY 125 -----RVOKSKISLYKKTWESKAYDQCRDADDAEQAEFVSANG--HQKV 170
DB 126 FRESRAEDGFRKAKPMLKRLKEVEASKSYHAARKDEKTAQTRSHAKADSAYSQEO 185
QY 171 EKSOMKAKCKESATEAEERYRQNIQLEERATEMEQEHRTTCEAFQLOEFDRLTLRNA 230
DB 186 RKLDERVERCAEAKETKAYQYEQTLAEHRYTPRWEDDEQAFTCOAERORLFEKDM 245
QY 231 LNVHCNQLSM-OCVADDELYEEVRLTLECGDVEGDIINGFIQSKSTGREPPAPVQNYD 289
DB 246 LTLTHQHLDSLSEKFEHLHRLHOGIEASDEEDLRWRRSTHGPAMANNVQFEKMSD 305
QY 290 REVTPPLIG-----SPSIOPSGVYIKRPSGLLHSGPKTTPSPAPASTETLTPTPRNEL 342
DB 306 TQRT--ISRKKEGKSPDEVLTLSIVPTBDGT--APPOSFSGPGQDEWSEDESPRK 361
QY 343 VYASIEVGATQGNLSSAQVRYALDYTPAQNDSDELISAG--DLNAVILEGEGMWVE-R 400
DB 362 AATGVR-----RALDYAGQEDDELFRAGEELLKMSSEDEGWCQGLQ 407
QY 401 NGORGFVPGSYLE 413
DB 408 SGRIGLYPANYVE 420
RESULT 11
PAC1_RAT STANDARD; PRT; 441 AA.
AC 09ZOM5:
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein kinase C and casein kinase substrate in neurons protein 1
DE (Synaptic, dynamin-associated protein I) (Synadin 1) (Synadin-1
DE (Sdpl) (Dynamin proline-rich domain-interacting protein) (Dynamin PBD-
DE interacting protein).
CN PACSIN1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 71-90 AND 412-429.
RP STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=99135893; PubMed=9950691;
RA Qualmann B., Roos J., Digregorio P.J., Kelly R.B.
RT "Synadin I, a synaptic dynamin-binding protein that associates with
the neural Wiskott-Aldrich syndrome protein.";
RL Mol. Biol. Cell 10:501-513(1999).
CC -1- FUNCTION: May play a role in vesicle formation and transport.
CC -1- SUBUNIT: Homo- and hetero-aggregates with other PACSINs. Binds
CC dynamin I, synaptojanin, synapsin I and the neural Wiskott-Aldrich
CC syndrome protein (N-WASP).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic. Colocalizes with dynamin I at
CC vesicular structures in the cell body and neurites.
CC -1- TISSUE SPECIFICITY: Highly expressed in brain.

CC -1- PTM: Phosphorylated by casein kinase 2 (CK2) and protein kinase C
CC (PKC) (Probable).
CC -1- SIMILARITY: BELONGS TO THE PACSIN FAMILY.
CC -1- SIMILARITY: Contains 1 FCH domain.
CC -1- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL collaboration
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF104402; AAD16887.1; .
CC HSSP: Q60631; 1CBO.
DR InterPro: IPR001060; Cdc15_Fes_C1P4.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00611; FCH; 1.
DR Pfam: PF00018; SH3; 1.
DR PRINTS: PR00452; SH3DOMAIN.
DR ProDom: PD000066; SH3; 1.
DR SMART: SM00055; FCH; 1.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS50133; FCH; 1.
DR PROSITE: PS50002; SH3; 1.
KW Endocytosis; SH3 domain; Coiled coil; Phosphorylation.
FT DOMAIN 10 73 FCH.
FT DOMAIN 382 441 SH3.
FT DOMAIN 144 165 COILED COIL (POTENTIAL).
FT DOMAIN 183 217 COILED COIL (POTENTIAL).
SQ SEQUENCE 441 AA: 50449 MW: 585B32850572DF8 CRC64:
Query Match 12.4%; Score 266.5; DB 1; Length 441;
Best Local Similarity 21.3%; Pred. No. 1,2e-08;
Matches 99; Conservative 82; Mismatches 202; Indels 81; Gaps 15;

QY 3 AQLQFADAEWCRDFTAHGTG-YEVLLQRLDGRKMKVVEELLNORAOAEERYGELYOIA 61
DB 8 ASEELTDSFW-----EVGNKRTVKRIDGHRCLNDIMSCYDERAKIEKAYAOQLTDWA 61
QY 62 RK-----AGGQEMNSLRTSPDLSKQOTENGVSAHIQIALALREELSLSEFFEROKE- 114
DB 115 -----QKKYEAIMRVOKSKISLYKKTWESKAYDQCRDADDA--EQAEFRVS 162
DB 118 YHKQINGEKEKTEAEADGFRKAKPVAKKYKLEAKKAYHACKKEKLTAMREMSKTE 177
QY 163 ANGHQKQVESOKKAKCKESATEAEERYRQNIQLEERATEMEQEHRTTCEAFQLOEFD 222
DB 178 QSTPPDQKKLVKVDKCRDQVOKTQEKYKVLVDGKTPQTMWEGHEQYFEGCQCPREK 237
QY 223 RLTLIR-----NALVHCNQLSMQCVADDELYEEVRLTLECGDVEGDIINGF 268
DB 238 RLVLKEVLVDLRHLNLAMENSSYIH-----VRELEQAIRGADAQEDLRWF 284
QY 269 IQSKSTGREPPAPVQNYTIDREVTPLTGSPSIOPSGVYIKRPSGLLHSGPKTTPSPAPAA 328
DB 285 RSTSGPGM--PMNWPFEEENPDLPHYAKKEKQPK---KAGGAALSNATGAVESTSOA 338
QY 329 STELTLPPTPERNELVYASIEVQATQGN-----LNSAODY-----RALYDYTAON 373
DB 339 GDSGVSSYIDRQO-AYATEMSDESGNPFEGCANANGANFEDDACKVRRALYDIDGOE 397
QY 374 SDELDISAGDILAVILEGEDGWTVER--NGORGFVPGSYLEK 415
DB 398 ODELSFKAGDELTKLGEDEGRCGRRLDGGQLGTPANYVEAI 441
RESULT 12
PAC2_CHICK STANDARD; PRT; 448 AA.

DR PROSITE: PS50133; FCH; 1.
 DR PROSITE: PS50002; SH3; 1.
 KM Endocytosis; SH3 domain; Coiled coil; Phosphorylation.
 FT DOMAIN 11 FCH.
 FT DOMAIN 417 477 SH3.
 FT DOMAIN 184 218 COILED COIL (POTENTIAL).
 SO SEQUENCE 477 AA; 55148 MW; FA6D187769EF9383 CRC64;
 Query Match 12.08; Score 259; DB 1; Length 477;
 Best Local Similarity 21.28; Pred. No. 4,4e-08;
 Matches 100; Conservative 88; Mismatches 208; Indels 76; Gaps 14;
 QY 9 DAFWCRDPTAHGTG-YEVLLQRLDGRKCKDVEELLQROAERPKELVQIARKKGGQ 67
 16 DSWF-----EVGKRTVAKRIDGGHLCNDLMCHERRARIEVVAQQLTEWAKRKQOL 69
 QY 68 TE-----MNSLRTPSDSKQOTENVGSAHQLALALREELSLEEFERORE----- 114
 70 VERGPQYGVTEKAMHNLMTAEKYSLEHLVKNALNNE--DFEIKKMWQKEAFHKQMMGG 127
 QY 115 --ORRKYEAIMDRVOKSKLSLYKKTMESKRAYQCKDAD--DAEQAFERVSANGHOKY 170
 128 FKRTKADDFGFRKAQCPWAKKLEVERAKKSTHAACEEKLATSRRTNSRADPAMNEQL 187
 QY 171 EKSQNAKQCKESATAEARYVRONIDQLERARTWEEQEHRTTCEAFQLOEFDRLLILRNA 230
 188 KRLQDKVEKSKQSQKTEKREKESLMDLDTPTQYEMNMQVFEQCCQFEDKRLSPFREY 247
 QY 231 LWHVHQMOLSMQCKD-DELYEYRLILEGCDVSGDINGFQSKSTGEPAPVYONYD 289
 248 LLEVEKHLDLINSEVASYIRELEYAIAKSDAMEDLKWPFNNHGPQSMNP-QFEDWSA 306
 QY 290 REVYPLIGSPSIDPSCGVIKRFSGL-----LHGSPTKTPSPAPAST--ETLLPT 336
 307 DLNRLTSREKKKPTGTV--TLTGISQSGROSSIQNHSHLSVQASQSTNPFDEDEET 364
 QY 337 PERNELVYASIE-----VQATQGNLNSAOD-----YRA 365
 365 VSLNENENKRIENGVSEKTHPAEMSDDESNNPFNSDFNGDNPPDEDLTLLEVVR 424
 QY 366 LYDYTPQNSDELISAGDILAVILEGEDGWTVER--NGCRGVPSYLEKL 415
 425 LVDYDQGEIDELSFKAGEELTIEDDEQGWCKGRLEGQGVGLYANVESV 476
 Db
 RESULT 15
 GAST_HUMAN STANDARD; PRT; 412 AA.
 AC 060861: 043144;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Growth-arrest-specific protein 7 (GAS-7).
 GN GAS7 OR KIA00394.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Brain;
 RX MEDLINE=9811655; PubMed=9455477;
 RA Ishikawa K.-I., Nagase T., Nakajima D., Seki N., Ohira M.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. VIII.
 RT 78 new cDNA clones from brain which code for large proteins in
 RT vitro";
 RL DNA Res. 4:307-313(1997).
 RP [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Cerebellum;
 RA Chao C.C.-K., Kuo J.-T., Ju Y.-T., Lin-Chao S.;
 RT "Human gas7: conservation of a mammalian growth-arrest-specific

RT gene";
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Eye, and Skin;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marsina K., Farmer A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., USCIN T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mulhally S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bonifield G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [4]
 RP CHROMOSOMAL TRANSLOCATION.
 RX MEDLINE=20183971; PubMed=10706619;
 RA Megonigal M.D., Cheung N.-K.V., Rappaport E.F., Nowell P.C.,
 RA Williams R.B., Jones D.H., Addya K., Leonard D.G.B., Kushner B.H.,
 RA Williams T.M., Lange B.J., Felix C.A.;
 RT "Detection of leukemia-associated MLL-GAS7 translocation early during
 RT chemotherapy with DNA topoisomerase II inhibitors";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:2814-2819(2000).
 CC -1- FUNCTION: MAY PLAY A ROLE IN PROMOTING MATURATION AND
 CC MORPHOLOGICAL DIFFERENTIATION OF CEREBELLAR NEURONS.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=O60861-1; Sequence=Displayed;
 CC Name=2; Synonyms=GAS7-CB;
 CC IsoId=O60861-2; Sequence=VSP_006802; VSP_006803;
 CC -1- DISEASE: INVOLVED IN ACUTE MYELOID LEUKEMIA BY A CHROMOSOMAL
 CC TRANSLOCATION T(11;17)(q23;p13) THAT INVOLVES MLL AND GAS7.
 CC -1- SIMILARITY: Contains 1 FCH domain.
 CC -1- SIMILARITY: Contains 1 WW domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

 DR EMBL: AB007854; BAA23690.1; -
 DR EMBL: AF224876; CAA12177.1; -
 DR EMBL: AF231998; AAG26332.2; ALT_INIT.
 DR EMBL: BC001152; AAH01152.1; -
 DR EMBL: BC006454; AAH06454.1; -
 DR HSSP: Q13526; IPIN.
 DR GeneID: HGNC:4169; GAS7.
 DR MIM: 603127; -
 DR GO: GO:0003700; F:transcription factor activity; TAS.
 DR GO: GO:0007050; P:cell cycle arrest; TAS.
 DR InterPro: IPR001060; Cdc15_Fes_CIP4.
 DR InterPro: IPR001202; WW_Rsp5_WWP.
 DR Pfam: PF00611; FCH; 1.
 DR Pfam: PF00397; WW; 1.
 DR SMART: SM00055; FCH; 1.
 DR SMART: SM00456; WW; 1.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2003, 09:15:23 ; Search time 20 Seconds

(without alignments)
877.950 Million cell updates/sec

Title: US-09-068-377c-1

Perfect score: 2152

Sequence: 1 MAAOLQFRDAFWCRDFTAHY.....WTVERNGRGVPGSYLEKL 415

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/6CTUS_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/6ackfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	2152	100.0	415	3	US-08-938-830-1
2	2152	100.0	415	3	US-09-020-222-1
3	1913.5	88.9	416	4	US-09-006-428A-1
4	1913.5	88.9	416	4	US-09-006-428A-17
5	1748	81.2	397	4	US-09-006-428A-2
6	1748	81.2	397	4	US-09-006-428A-19
7	1714.5	79.7	400	3	US-08-938-830-29
8	789	36.7	166	4	US-09-006-428A-3
9	325	15.1	907	3	US-08-938-830-26
10	325	15.1	907	3	US-09-020-222-26
11	281	13.1	57	4	US-09-006-428A-5
12	265.5	12.3	441	4	US-08-630-915A-34
13	246	11.4	48	3	US-08-938-830-3
14	246	11.4	48	3	US-09-020-222-3
15	215.5	10.0	162	4	US-09-006-428A-4
16	197	9.2	377	4	US-08-630-915A-36
17	181	8.4	788	4	US-08-630-915A-30
18	179.5	8.3	433	4	US-08-630-915A-18
19	169.5	7.9	434	4	US-08-630-915A-22
20	168.5	7.8	546	2	US-08-942-423-4
21	167	7.8	546	4	US-08-630-915A-14
22	161	7.5	1151	3	US-08-630-915A-6
23	161	7.5	1200	3	US-08-840-006-5
24	154.5	7.2	182	4	US-08-630-915A-200
25	153	7.1	355	4	US-08-630-915A-192
26	150.5	7.0	451	1	US-08-435-454-4
27	150.5	7.0	451	2	US-08-919-145-6

28	150.5	7.0	451	3	US-09-344-889-6	Sequence 6, Appl
29	150.5	7.0	451	5	PCT-US96-06231A-4	Sequence 4, Appl
30	149.5	6.9	451	1	US-08-652-972A-4	Sequence 4, Appl
31	149.5	6.9	451	3	US-08-870-126-4	Sequence 4, Appl
32	149.5	6.9	451	4	US-09-445-247-4	Sequence 4, Appl
33	148.5	6.9	509	4	US-08-630-915A-194	Sequence 194, App
34	148.5	6.9	534	2	US-08-878-563A-1	Sequence 1, Appl
35	148.5	6.9	534	3	US-09-270-117-1	Sequence 1, Appl
36	147.5	6.9	2101	1	US-08-466-390-4	Sequence 4, Appl
37	147.5	6.9	2101	1	US-08-470-950-4	Sequence 4, Appl
38	147.5	6.9	2101	1	US-08-467-781-4	Sequence 4, Appl
39	147.5	6.9	2101	1	US-08-195-487-4	Sequence 4, Appl
40	147.5	6.9	2101	2	US-08-483-924-4	Sequence 4, Appl
41	147.5	6.9	2101	3	US-09-452-294-1	Sequence 1, Appl
42	147.5	6.9	2101	5	PCT-US93-06160-4	Sequence 4, Appl
43	146	6.8	521	2	US-08-878-563A-3	Sequence 3, Appl
44	146	6.8	521	3	US-09-270-117-3	Sequence 3, Appl
45	146	6.8	605	4	US-08-714-741-46	Sequence 46, Appl

ALIGNMENTS

```
RESULT 1
US-08-938-830-1
; Sequence 1, Application US/08938830
; Patent No. 6040437
; GENERAL INFORMATION:
; APPLICANT: Lasky, Laurence A.
; TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage
; TITLE OF INVENTION: Furrow-Associated Proteins (PSTRIPs)
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 MB floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,830
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/796419
; FILING DATE: 07-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: P1066P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/952-9881
; TELEFAX: 650/225-3216
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 415 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; US-08-938-830-1

Query Match 100.0%; Score 2152; DB 3; length 415;
Best Local Similarity 100.0%; Pred. No. 1.2e+189;
Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAOLQFRDAFWCRDFTAHYEVLLQRLDGRKMKDVEELLQRAQAERYGKELVQI 60
Db 1 MAAOLQFRDAFWCRDFTAHYEVLLQRLDGRKMKDVEELLQRAQAERYGKELVQI 60
```

QY 61 ARKAGGOTEMNSLRTSPDSLKQOTENVGSAHIQIALALREELRSLEEFREKOEORRKYE 120
DB 61 ARKAGGOTEMNSLRTSPDSLKQOTENVGSAHIQIALALREELRSLEEFREKOEORRKYE 120
QY 121 AIMRVQKSKLSLYKKTMEKSKKAYDQKCRDADDAEQAFERVSANGHOKOVEKSONKAKOC 180
DB 121 AIMRVQKSKLSLYKKTMEKSKKAYDQKCRDADDAEQAFERVSANGHOKOVEKSONKAKOC 180
QY 181 KESATEAEVRYRONIEOLERARTEWEOEHRTTCEAFQLOEFDRLLTLRNALMVHONOLSM 240
DB 181 KESATEAEVRYRONIEOLERARTEWEOEHRTTCEAFQLOEFDRLLTLRNALMVHONOLSM 240
QY 241 QCVKDELEYEVRLLTEGCDVGEINGFIQSKSTGREPPAPVPYQNYDYREVTPPLIGSPS 300
DB 241 QCVKDELEYEVRLLTEGCDVGEINGFIQSKSTGREPPAPVPYQNYDYREVTPPLIGSPS 300
QY 301 IOPSCGVIKRFSGLHSGPKTTPSAPASTETLTPTPERNELVYASIEVOATQGNLNSA 360
DB 301 IOPSCGVIKRFSGLHSGPKTTPSAPASTETLTPTPERNELVYASIEVOATQGNLNSA 360
QY 361 QDYRALYDTAONSDDELISAGDILAVILLEGEDGWTVERNGQRGFVPGSYLEKL 415
DB 361 QDYRALYDTAONSDDELISAGDILAVILLEGEDGWTVERNGQRGFVPGSYLEKL 415

RESULT 2

US-09-020-222-1

Sequence 1, Application US/09020222

Patent No. 6111073

GENERAL INFORMATION:

APPLICANT: Lasky, Laurence A.

TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage

NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

STREET: 1 DNA Way

CITY: South San Francisco

STATE: California

COUNTRY: USA

ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WinPatlin (Genentech)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/020,222

FILING DATE: 06-Feb-1998

CLASSIFICATION:

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: 08/798419

FILING DATE: 02/07/1997

ATTORNEY/AGENT INFORMATION:

NAME: Dreger, Ginger R.

REGISTRATION NUMBER: 33,055

REFERENCE/DOCKET NUMBER: P1066r1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650/225-3216

TELEFAX: 650/952-9881

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 415 amino acids

TYPE: Amino Acid

TOPOLOGY: Linear

US-09-020-222-1

Query Match 100.0%; Score 2152; DB 3; Length 415;
Best Local Similarity 100.0%; Pred. No. 1.2e-189;
Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MMAQLOFRAFWCRDFTAHGTGVEVLLQRLIDGRKMKDVEELLRQRAQAEERIGKELVQI 60

DB 1 MMAQLOFRAFWCRDFTAHGTGVEVLLQRLIDGRKMKDVEELLRQRAQAEERIGKELVQI 60
QY 61 ARKAGGOTEMNSLRTSPDSLKQOTENVGSAHIQIALALREELRSLEEFREKOEORRKYE 120
DB 61 ARKAGGOTEMNSLRTSPDSLKQOTENVGSAHIQIALALREELRSLEEFREKOEORRKYE 120
QY 121 AIMRVQKSKLSLYKKTMEKSKKAYDQKCRDADDAEQAFERVSANGHOKOVEKSONKAKOC 180
DB 121 AIMRVQKSKLSLYKKTMEKSKKAYDQKCRDADDAEQAFERVSANGHOKOVEKSONKAKOC 180
QY 181 KESATEAEVRYRONIEOLERARTEWEOEHRTTCEAFQLOEFDRLLTLRNALMVHONOLSM 240
DB 181 KESATEAEVRYRONIEOLERARTEWEOEHRTTCEAFQLOEFDRLLTLRNALMVHONOLSM 240
QY 241 QCVKDELEYEVRLLTEGCDVGEINGFIQSKSTGREPPAPVPYQNYDYREVTPPLIGSPS 300
DB 241 QCVKDELEYEVRLLTEGCDVGEINGFIQSKSTGREPPAPVPYQNYDYREVTPPLIGSPS 300
QY 301 IOPSCGVIKRFSGLHSGPKTTPSAPASTETLTPTPERNELVYASIEVOATQGNLNSA 360
DB 301 IOPSCGVIKRFSGLHSGPKTTPSAPASTETLTPTPERNELVYASIEVOATQGNLNSA 360
QY 361 QDYRALYDTAONSDDELISAGDILAVILLEGEDGWTVERNGQRGFVPGSYLEKL 415
DB 361 QDYRALYDTAONSDDELISAGDILAVILLEGEDGWTVERNGQRGFVPGSYLEKL 415

RESULT 3

US-09-006-428A-1

Sequence 1, Application US/09006428A

Patent No. 644439

GENERAL INFORMATION:

APPLICANT: Kazuhisa Nishizawa

APPLICANT: Wenqian An

APPLICANT: Ellis L. Reinherz

TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF A

FILE REFERENCE: 1062.1020-000

CURRENT APPLICATION NUMBER: US/09/006,428A

CURRENT FILING DATE: 1998-01-13

NUMBER OF SEQ ID NOS: 28

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 1

LENGTH: 416

TYPE: PRT

ORGANISM: Homo sapiens

US-09-006-428A-1

Query Match 88.9%; Score 1913.5; DB 4; Length 416;
Best Local Similarity 88.0%; Pred. No. 1.1e-167;
Matches 366; Conservative 20; Mismatches 29; Indels 1; Gaps 1;

DB 1 MMAQLOFRAFWCRDFTAHGTGVEVLLQRLIDGRKMKDVEELLRQRAQAEERIGKELVQI 60

```

Oy      301  |IOPSGGVTKRRPSGLIHGSPKTP-SAPAASTETLPTPERNLVYASIEVQATOGNLNSS 355
Db      301  |IOPSGGMKRRPSGLIHGSPKTPSLAASASTETLPTPERNGVYTAIAVQETIQNPASP 360
Oy      360  |AODYRALDYTAQNSDELDISAGDILAVILEGEDGMMVYERNRGQGFVPGSYLEKL 415
         |||||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      361  |AQERFALDYTAQNDELDLSAGDILLEVILEBEDGMMVYERNRGQGFVPGSYLEKL 416

RESULT 4
US-09-006-428A-17
: Sequence 17, Application US/09006428A
: Patent No. 6444439
: GENERAL INFORMATION:
:   APPLICANT: Jing Li
:   APPLICANT: Kazuhisa Nishizawa
:   APPLICANT: Wenqian An
:   APPLICANT: Ellis L. Reinherz
:   TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF A
:   FILE REFERENCE: Gdcl5-LIKE ADAPTOR PROTEIN (CD2BP1)
:   CURRENT APPLICATION NUMBER: US/09/006,428A
:   CURRENT FILING DATE: 1998-01-13
:   NUMBER OF SEQ ID NOS: 28
:   SOFTWARE: FastSeq for Windows Version 4.0
:   SEQ ID NO 17
:   LENGTH: 416
:   TYPE: PRT
:   ORGANISM: Homo sapiens
: US-09-006-428A-17

```

RESULT 5
 US-09-006-428A-2
 ; Sequence 2, Application US/09006428A
 ; Patent No. 6444439
 ; GENERAL INFORMATION:
 ; APPLICANT: Jing Li
 ; APPLICANT: Kazuhisa Nishizawa
 ; APPLICANT: Wenqian An

[illegible]

```

: APPLICANT: Ellis L. Reinherz
: TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF A
: TITLE OF INVENTION: CDG15-LIKE ADAPTOR PROTEIN (CD2BP1)
: FILE REFERENCE: 1062.1020-000
: CURRENT APPLICATION NUMBER: US/09/006.428A
: CURRENT FILING DATE: 1998-01-13
: NUMBER OF SEQ. ID NOS: 28
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2
: LENGTH: 397
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-006-428A-2

```

	Query Match	81.2%; Score 1748; DB 4; Length 397;
	Best Local Similarity	82.0%; Pred. No. 1.7e-15;
	Matches	Conservative 20; Mismatches 35; Indels 20; Gaps 2.
QY	1 MMAQLQERDAPWCFKDFTAHTGYEVLRLDLDRGKKCKVDEELRLROQAQAEERYGKEIYOI	60
Dd	1 MMPQLQFRDAPWCDFTAHTGYEVLLRLLDRGKKCKMEELLRQQAEEERYGKEIYOI	60
QY	61 ARKAGGGOTEMNSLRTSPDSLKOQTENNGSAHITOLALAREELRSLEEFEREOKEQRKYE	120
Dd	61 ARKAGGGOTEINLSRASPSFSLTKOMENWGSSHIQTLALTREELRSLEEFEREOKEQRKYE	120
QY	121 AIMRVQRKSKLSLYKTTHESKAARDQCRDDADDAQAFERYSANGHOKOVEKSOMKACQ	180
Dd	121 AVMRVQRKSKLSLYKKAKMESKRYEQKCRDDADDAQAFERISANGHOKOVEKSOMKACQ	180
QY	181 KESATVEARVRVROMIEOLERTAMEOEHRPTTCEAFOLOQEEDRLITLENALMVHCNOISM	240
Dd	181 KDSATEARVRVROSIAOLEKRAEMEDPHRTTCFAFOLOQEDRLITLENALMVHSNOLISM	240
QY	241 QCVKDELLEYEVRRLTECGDVEGDINGFIQSSTGRREPAPVPYONYNDREVTPILIGSPS	300
Dd	241 QCVADELLEYEVRRLTECCSIDADIDISTQAKSTGEPEPGEVRLAD-----	286
QY	301 IQPSCGVIKRPSSLHGSPKTT-P-SAPASETLTLPFERNELVYASILEVOATOGNLMS	359
Dd	287 -----SASRSPSGLLHGSPKTTSLAASAASETLPPPERNEGVTATAVQEIQGNPASP	341
QY	360 AQDYRALYDYAONSDDELISAGDIIVYLLEGEGMWTVERNNGOGEFPAGSYLEKL	415
Dd	342 AQETRALYDYAQNPDDELISAGDILEVYLEGEGMWTVERNGORGEPFGSYLEKL	397

RESULT 6
US-09-006-428A-19

- / Sequence 19, Application US/09006428A
- / Patent No. 6444439
- / GENERAL INFORMATION:
- / APPLICANT: Jing Li
- / APPLICANT: Kazuhisa Nishizawa
- / APPLICANT: Wengian An
- / TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF A
- / FILE REFERENCE: cdc15-LIKE ADAPTOR PROTEIN (CD2BP1)
- / CURRENT FILING DATE: 1998-01-13
- / NUMBER OF SEQ ID NOS: 28
- / SOFTWARE: FastSeq for Windows Version 4.0
- / SEQ ID NO 19
- / LENGTH: 397
- / TYPE: PRT
- / ORGANISM: Homo sapiens
- / US-09-006-428A-19

Query Match	81.2%;	Score 1748;	DB 4;	length 397;
Best Local Similarity	82.0%;	Pred. No. 1.7e-152;		
Matches 341;	Conservative 20;	Mismatches 35;	Indels 20;	Gaps 2;

```

OY 1 MAAQLQFDPAFCRDTAHTGTEVLLQRLLDGRKKMKDVEELLRQRAQAEERKGLVQI 60
  1 MAAQLQFDPAFCRDTAHTGTEVLLQRLLDGRKKMKDVEELLRQRAQAEERKGLVQI 60
DB 1 MAAQLQFDPAFCRDTAHTGTEVLLQRLLDGRKKMKDVEELLRQRAQAEERKGLVQI 60
OY 61 ARKAGGQTEMSLRTSFDSLKQOOTENVSASHIOLALALREELSLSEFFRERQEKORRKE 120
  61 ARKAGGQTEMSLRTSFDSLKQOOTENVSASHIOLALALREELSLSEFFRERQEKORRKE 120
DB 61 ARKAGGQTEMSLRTSFDSLKQOOTENVSASHIOLALALREELSLSEFFRERQEKORRKE 120
OY 121 AIMRVQSKSLTYKKTMSKAKAYDQKCDADDAQAFPRVANSANGQKQVEKSNKAKQC 180
  121 AIMRVQSKSLTYKKTMSKAKAYDQKCDADDAQAFPRVANSANGQKQVEKSNKAKQC 180
DB 121 AIMRVQSKSLTYKKTMSKAKAYDQKCDADDAQAFPRVANSANGQKQVEKSNKAKQC 180
OY 181 KESATEARVYRONIEQLERARTEWEOEHRTTCEAFQLOEFDRLLTILRNALMVHNCQLSM 240
  181 KESATEARVYRONIEQLERARTEWEOEHRTTCEAFQLOEFDRLLTILRNALMVHNCQLSM 240
DB 181 KESATEARVYRONIEQLERARTEWEOEHRTTCEAFQLOEFDRLLTILRNALMVHNCQLSM 240
OY 241 QCVKDELEYEVRLLTEGCGVDGDIINGFIQSKSTGCRPPAPVYQNYQREVYTPPLIGSPS 300
  241 QCVKDELEYEVRLLTEGCGVDGDIINGFIQSKSTGCRPPAPVYQNYQREVYTPPLIGSPS 300
DB 241 QCVKDELEYEVRLLTEGCGVDGDIINGFIQSKSTGCRPPAPVYQNYQREVYTPPLIGSPS 300
OY 301 IOPSGVYIKRPSGLHSGPKTTP-SAPASTETLTPPERNELVYASIEVQATQGNLSS 359
  301 IOPSGVYIKRPSGLHSGPKTTP-SAPASTETLTPPERNELVYASIEVQATQGNLSS 359
DB 287 -----SASRSRGLHSGPKTTLASAASASTETLTPPERNCGVYTAIAVQELQGNPASP 341
  287 -----SASRSRGLHSGPKTTLASAASASTETLTPPERNCGVYTAIAVQELQGNPASP 341
OY 360 AODYRALYDYTAQNSDELDISAGDILAVILLEGEDGQWTVERNQGRGVPGSTLEKL 415
  360 AODYRALYDYTAQNSDELDISAGDILAVILLEGEDGQWTVERNQGRGVPGSTLEKL 415
DB 342 AOERYALDYTAQNDELDELISAGDILEVILLEGEDGQWTVERNQGRGVPGSTLEKL 397
  342 AOERYALDYTAQNDELDELISAGDILEVILLEGEDGQWTVERNQGRGVPGSTLEKL 397

```

RESULT 7
US-08-938-830-29
Sequence 29, Application US/08938830
Patent No. 6040437

GENERAL INFORMATION:
APPLICANT: Lasky, Laurence A.
TITLE OF INVENTION: Tyrosine phosphorylated Cleavage
TITLE OF INVENTION: Furorow-Associated Proteins (PSTPIPS)
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpath (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938,830
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/798419
FILING DATE: 07-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
TELECOMMUNICATION INFORMATION:
REFERENCE/DOCKET NUMBER: P1066P1
TELEPHONE: 650/225-3216
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 400 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-938-830-29

Query Match 79.7%; Score 1714.5; DB 3; Length 400;

Best Local Similarity 77.9%; Pred. No. 2e-149;
Matches 338; Conservative 17; Mismatches 26; Indels 53; Gaps 3;
OY 1 MAAQLQFDPAFCRDTAHTGTEVLLQRLLDGRKKMKDVEELLRQRAQAEERKGLVQI 60
 1 MAAQLQFDPAFCRDTAHTGTEVLLQRLLDGRKKMKDVEELLRQRAQAEERKGLVQI 60
DB 1 MAAQLQFDPAFCRDTAHTGTEVLLQRLLDGRKKMKDVEELLRQRAQAEERKGLVQI 60
OY 61 ARKAGGQTEMSLRTSFDSLKQOOTENVSASHIOLALALREELSLSEFFRERQEKORRKE 118
 61 ARKAGGQTEMSLRTSFDSLKQOOTENVSASHIOLALALREELSLSEFFRERQEKORRKE 118
DB 61 ARKAGGQTEMSLRTSFDSLKQOOTENVSASHIOLALALREELSLSEFFRERQEKORRKE 120
OY 119 -----YEAINDRVQSKSLTYKKTMSKAKAYDQKCDADDAQAFPRV 161
 119 -----YEAINDRVQSKSLTYKKTMSKAKAYDQKCDADDAQAFPRV 161
DB 121 AVPRQSDCMEVKSPEWEEYAVMDRVQSKSLTYKKTMSKAKAYDQKCDADDAQAFPRV 180
 121 AVPRQSDCMEVKSPEWEEYAVMDRVQSKSLTYKKTMSKAKAYDQKCDADDAQAFPRV 180
OY 162 SANGHOKOYKESQNKAKQCKESATEARVYRONIEQLERARTEWEOEHRTTCEAFQLOEF 221
 162 SANGHOKOYKESQNKAKQCKESATEARVYRONIEQLERARTEWEOEHRTTCEAFQLOEF 221
DB 181 SANGHOKOYKESQNKAKQCKESATEARVYRONIEQLERARTEWEOEHRTTCEAFQLOEF 240
 181 SANGHOKOYKESQNKAKQCKESATEARVYRONIEQLERARTEWEOEHRTTCEAFQLOEF 240
OY 222 DRLLTILRNALMVHNCQLSMOCVMDDELEYEVRLLTEGCGVDGDIINGFIQSKSTGCRPPAP 281
 222 DRLLTILRNALMVHNCQLSMOCVMDDELEYEVRLLTEGCGVDGDIINGFIQSKSTGCRPPAP 281
DB 241 DRLLTILRNALMVHNCQLSMOCVMDDELEYEVRLLTEGCGVDGDIINGFIQSKSTGCRPPAP 298
 241 DRLLTILRNALMVHNCQLSMOCVMDDELEYEVRLLTEGCGVDGDIINGFIQSKSTGCRPPAP 298
OY 282 VPYQNYDREVPPLIGSPSIQPSIQRPSGLHSGPKTTPSAPASTETLTPPERNE 341
 282 VPYQNYDREVPPLIGSPSIQPSIQRPSGLHSGPKTTPSAPASTETLTPPERNE 341
DB 299 -----RFGSLHSGPKTTPSASAGSTETLTPPERNE 330
 299 -----RFGSLHSGPKTTPSASAGSTETLTPPERNE 330
OY 342 IYVASTEQATQGNLSSADYRALYDYTAQNSDELDISAGDILAVILLEGEDGQWTVERN 401
 342 IYVASTEQATQGNLSSADYRALYDYTAQNSDELDISAGDILAVILLEGEDGQWTVERN 401
DB 331 GVTALAVQELQGNPASPADYRALYDYTAQNSDELDISAGDILAVILLEGEDGQWTVERN 386
 331 GVTALAVQELQGNPASPADYRALYDYTAQNSDELDISAGDILAVILLEGEDGQWTVERN 386
OY 402 GQGFVPGSTLEKL 415
 402 GQGFVPGSTLEKL 415
DB 387 GQGFVPGSTLEKL 400
 387 GQGFVPGSTLEKL 400

RESULT 8
US-09-006-428A-3
Sequence 3, Application US/09006428A
Patent No. 644439
GENERAL INFORMATION:
APPLICANT: Jiong Li
APPLICANT: Kazuhisa Nishizawa
APPLICANT: Wenqian An
APPLICANT: Ellis L. Reinherz
TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF A
FILE REFERENCE: 1062.1020-000
CURRENT APPLICATION NUMBER: US/09/006,428A
CURRENT FILING DATE: 1998-01-13
NUMBER OF SEQ ID NOS: 28
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 166
TYPE: PRT
ORGANISM: Homo sapiens
US-09-006-428A-3

Query Match 36.7%; Score 789; DB 4; Length 166;
Best Local Similarity 88.0%; Pred. No. 7.3e-65;
Matches 146; Conservative 10; Mismatches 10; Indels 0; Gaps 0;
OY 123 MDRVQSKSLTYKKTMSKAKAYDQKCDADDAQAFPRVANSANGHOKOYKESQNKAKQCKE 182
 123 MDRVQSKSLTYKKTMSKAKAYDQKCDADDAQAFPRVANSANGHOKOYKESQNKAKQCKE 182
DB 1 MDRVQSKSLTYKKTMSKAKAYDQKCDADDAQAFPRVANSANGHOKOYKESQNKAKQCKE 60
 1 MDRVQSKSLTYKKTMSKAKAYDQKCDADDAQAFPRVANSANGHOKOYKESQNKAKQCKE 60
OY 183 SATEARVYRONIEQLERARTEWEOEHRTTCEAFQLOEFDRLLTILRNALMVHNCQLSMOC 242
 183 SATEARVYRONIEQLERARTEWEOEHRTTCEAFQLOEFDRLLTILRNALMVHNCQLSMOC 242
DB 61 SATEARVYRONIEQLERARTEWEOEHRTTCEAFQLOEFDRLLTILRNALMVHNCQLSMOC 120
 61 SATEARVYRONIEQLERARTEWEOEHRTTCEAFQLOEFDRLLTILRNALMVHNCQLSMOC 120
OY 243 VKDDELEYEVRLLTEGCGVDGDIINGFIQSKSTGCRPPAPVYQNY 288
 243 VKDDELEYEVRLLTEGCGVDGDIINGFIQSKSTGCRPPAPVYQNY 288
DB 121 VKDDELEYEVRLLTEGCGVDGDIINGFIQSKSTGCRPPAPVYQNY 166
 121 VKDDELEYEVRLLTEGCGVDGDIINGFIQSKSTGCRPPAPVYQNY 166

US-09-006-428A-5
; Sequence 5, Application US/09006428A
; Patent No. 644439
; GENERAL INFORMATION:
; APPLICANT: JUNG LI
; APPLICANT: Kazuhisa Nishizawa
; APPLICANT: Mengqian An
; APPLICANT: Ellis L. Reinherz
; TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF A
; FILE REFERENCE: 1062.1020-000
; CURRENT FILING DATE: 1998-01-13
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 57
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-006-428A-5

Query Match 13.18; Score 281; DB 4; Length 57;
Best Local Similarity 92.9%; Pred. No. 7, 5e-19;
Matches 52; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 360 AADYRALYDTAQNDELISAGDILAVILEGEGMWTVERNGORGFVPSYLEKL 415
DB 2 ADEYRALYDTAQNDELISAGDILAVILEGEGMWTVERNGORGFVPSYLEKL 57

RESULT 12
US-08-630-915A-34
; Sequence 34, Application US/08630915A
; Patent No. 630820
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: HOFFMAN, No. 6309820h
; APPLICANT: KAY, Brian K.
; APPLICANT: FOWLES, Dana M.
; APPLICANT: MCCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,915A
; FILING DATE: 03-APR-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Mastrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 441 amino acids
; TYPE: amino acid
; STRANDEDNESS:

TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-630-915A-34

Query Match 12.38; Score 265.5; DB 4; Length 441;
Best Local Similarity 21.2%; Pred. No. 4, 9e-16;
Matches 100; Conservative 84; Mismatches 192; Indels 95; Gaps 16;

QY 3 AGLFPAFCRCFTHTG-YEVLRLRLDKCMQDVEELLQRQAEERYKELVQTA 61
DB 8 ASEITDSEW-----EVGNKRTVRIDGHRCLNDLMSCVORAKIEATIAQQLTDA 61
QY 62 RK-----AGGQTEMNSLRTSPDLKQOTENWGAHIALALAREILRSLEPERROKE- 114
DB 62 KRWRLIEKGPQ--YSLERANGAMTENDAKVSELHQEVKNSLNE--DLEKXKNQKDA 117
QY 115 -----QKKYERLMDRVOKSKLSLYKKTMSKKAVDQCRDADA--EQAEFRS 162
DB 118 YHRQIMGFEKETAEEDGFERKAKPMAKKKLELAKKAYHLACKERLMTREMSKTE 177
QY 163 ANGHQKQVEKSNKAKQCKESATEAEKVYKONTQLEERATEMEQEHRTCEAFOLQEPD 222
DB 178 QSVTPEDQKKLVKVDKCRQDVOKTQETKEKVLQEDKTPQYMEGMEQVFECCQFEK 237
QY 223 RLTLRL-----NALMVHQNOLSMQCVKDDLEYEVRRLTLEGCDVEGDINGR 268
DB 238 RLVLFLKEVLDIRHNLNLAENSSYM-----YRLEGAIRGADAOEDLRWF 284
QY 269 IQSKSTGEPPAVPYONYIDREVTPLIGSPSIQPSGCVIKRPSGLIHGSPKTPSPAPAA 328
DB 285 RSTSGPGM--PMWMPQFEENWNPDLPHTTAKKERQ-----KKAEG-----ATLSNATG 330
QY 329 STELTPTPEPRLN-----VASTIEVQATQGN-----LNSADY-----RAL 366
DB 331 AVESTQAGRGVSYSYDRQTYATENSDESGNPGNANGANGANPEDDAGVRYRAL 390
QY 367 YDYTAQNSDELISAGDILAVILEGEGMWTVER--NGORGFVPSYLEKL 415
DB 391 YDYDGGQDELISFKADELTKLEDEQGCRCRRLDSGGGLVLPANVANI 441

RESULT 13
US-08-938-830-3
; Sequence 3, Application US/08938830
; Patent No. 604037
; GENERAL INFORMATION:
; APPLICANT: Lasky, Laurence A.
; APPLICANT: Dowbenko, Donald J.
; TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage
; TITLE OF INVENTION: Furrow-Associated Proteins (PSTPIPs)
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Winpatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,830
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/798419
; FILING DATE: 07-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055

REFERENCE/DOCKET NUMBER: P1066P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-3216
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 48 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-938-830-3

Query Match 11.4%; Score 246; DB 3; Length 48;
Best Local Similarity 97.9%; Pred. No. 9.6e-16;
Matches 46; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 366 LVDYTAQNSDELDISAGDILAVILLEGDGMWTVERNQGRGFVPSYL 412
DB 1 LVDYTAQNSDELDISAGDILAVILXGEGDGMWTVERNQGRGFVPSYL 47

RESULT 14
US-09-020-222-3
Sequence 3, Application US/09020222
Patent No. 6111073
GENERAL INFORMATION:
APPLICANT: Laskey, Laurence A.
TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage
TITLE OF INVENTION: Furrow-Associated Proteins (PSTRIPS)
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/020,222
FILING DATE: 06-Feb-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/798419
FILING DATE: 02/07/1997
ATTORNEY/AGENT INFORMATION:
NAME: Dreyer, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: P1066r1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-3216
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 48 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-09-020-222-3

Query Match 11.4%; Score 246; DB 3; Length 48;
Best Local Similarity 97.9%; Pred. No. 9.6e-16;
Matches 46; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 366 LVDYTAQNSDELDISAGDILAVILLEGDGMWTVERNQGRGFVPSYL 412
DB 1 LVDYTAQNSDELDISAGDILAVILXGEGDGMWTVERNQGRGFVPSYL 47

RESULT 15
US-09-006-428A-4

Sequence 4, Application US/09006428A
Patent No. 6444439
GENERAL INFORMATION:
APPLICANT: Jing Li
APPLICANT: Kazuhisa Nishizawa
APPLICANT: Wenqian An
APPLICANT: Ellis L. Reinherz
TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF A
TITLE OF INVENTION: CDG15-LIKE ADAPTOR PROTEIN (CD2BP1)
FILE REFERENCE: 1062.1020-000
CURRENT APPLICATION NUMBER: US/09/006,428A
CURRENT FILING DATE: 1998-01-13
NUMBER OF SEQ ID NOS: 28
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 162
TYPE: PRT
ORGANISM: Homo sapiens
US-09-006-428A-4

Query Match 10.0%; Score 215.5; DB 4; Length 162;
Best Local Similarity 30.1%; Pred. No. 4.1e-12;
Matches 50; Conservative 34; Mismatches 77; Indels 5; Gaps 2;

QY 123 MDRVQSKSLSLYKKTMSKKAYDQCRDADDAEQAFERVSANGHOKOYERSONKAKOCKE 182
DB 1 IEELYOKKTALEIDLESEKDAVEYSCKLNMYMOTKKMTG----REIDKYNLAKIROAAL 56
QY 183 SATEAERYRONIDOLEPARTMEOEHRTECEAFOLDFDLTLIRNALVHNCNLSMOC 242
DB 57 AVKMDAERYRETNELLVTRFEMIDRWTEVCDAFOHLEERYLFEFLKTMWMAVAMITISVAC 116
QY 243 VKDEIXEVRRLTEGCGVEDGINDGFTOSKSTGEPPAPRYONY 288
DB 117 VKDESECKIRLTLENTNIDEDITOMIONESTGTTIP-PLPEFNDY 161

Search completed: July 28, 2003, 09:17:53
Job time : 21 secs

THIS PAGE BLANK (USPTO)


```
QY 181 KESATAEAEVRYRONITOLEARTEMOEHRRTCEAFOLQOEFDRLTLIRNALVHNCQLSM 240
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 KDSATAEAEVRYROSIAOLEKVRAMEOEHRRTCEAFOLQOEFDRLTLIRNALVHNSQLSM 240
QY 241 QCVKDELYEVRITLLEGGDVEGDIINGFIQSKSTGPREPAVPYQNYVDREVPLTSSPG 300
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 241 QCVKDELYEVRITLLEGGDVEGDIINGFIQAKSTGTEPPAPVPYQNYVDREVPLTSSPG 300
QY 301 IOPSCGVIRKRFSGLLHSGSKTTP-SAPAASTETLTPPRNELVYASIEVQATOGNLNS 359
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 301 IOPSCGMIRKRFSGLLHSGSKTTP-SAPAASTETLTPPRNEGYVTAIAVQELQGNPASP 360
QY 360 AODYRALDYTAONSDLDISAGDILAVILLEGEDGMWTVERNQGRGFVGSYLEKL 415
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 361 AOEYRALDYTAONPDELDSLADGILLEVILLEGEDGMWTVERNQGRGFVGSYLEKL 416
```

```
RESULT 2
US-10-067-076-20
; Sequence 20, Application US/10067076
; Publication No. US20030104404A1
; GENERAL INFORMATION:
; APPLICANT: Wise, Carol A.
; TITLE OF INVENTION: Genetic Markers for Autoimmune Disorder
; FILE REFERENCE: TEX871/4-006US/36000
; CURRENT APPLICATION NUMBER: US/10/067,076
; PRIOR FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: 60/287,893
; PRIOR FILING DATE: 2001-05-01
; PRIOR APPLICATION NUMBER: 09/710,693
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Human Amino Acid
US-10-067-076-20
```

```
Query Match 88.7%; Score 1910.5; DB 15; Length 416;
Best Local Similarity 87.7%; Pred. No. 9.1e-146;
Matches 365; Conservative 21; Mismatches 29; Indels 1; Gaps 1;

QY 1 MMAQLOFRDAFWCRDFTAHGTGYEVLLQRLLDGRRKCKDVEELLRQRAQAEERYGKELVQI 60
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MMPOLQFKDAFWCRDFTAHGTGYEVLLQRLLDGRRKCKDVEELLRQRAQAEERYGKELVQI 60
QY 61 ARKAGGQTEINSLRASFDSLKQOETVNGSAHIQALALREBELSLLEEFREKQEKRYE 120
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 ARKAGGQTEINSLRASFDSLKQOETVNGSSHQIALALREBELSLLEEFREKQEKRYE 120
QY 121 AINDRVQKSKLSLYKTTMESKKAYDQKCRDADDAEQAFERVSANGHOKQVEKSNARQOC 180
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 AINDRVQKSKLSLYKTTMESKKAYDQKCRDADDAEQAFERISANGHOKQVEKSNARQOC 180
QY 181 KESATAEAEVRYRONITOLEARTEMOEHRRTCEAFOLQOEFDRLTLIRNALVHNCQLSM 240
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 KDSATAEAEVRYROSIAOLEKVRAMEOEHRRTCEAFOLQOEFDRLTLIRNALVHNSQLSM 240
QY 241 QCVKDELYEVRITLLEGGDVEGDIINGFIQSKSTGPREPAVPYQNYVDREVPLTSSPG 300
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 241 QCVKDELYEVRITLLEGGDVEGDIINGFIQAKSTGTEPPAPVPYQNYVDREVPLTSSPG 300
QY 301 IOPSCGVIRKRFSGLLHSGSKTTP-SAPAASTETLTPPRNELVYASIEVQATOGNLNS 359
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 301 IOPSCGMIRKRFSGLLHSGSKTTP-SAPAASTETLTPPRNEGYVTAIAVQELQGNPASP 360
QY 360 AODYRALDYTAONSDLDISAGDILAVILLEGEDGMWTVERNQGRGFVGSYLEKL 415
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 361 AOEYRALDYTAONPDELDSLADGILLEVILLEGEDGMWTVERNQGRGFVGSYLEKL 416
```

```
US-10-067-076-22
; Sequence 22, Application US/10067076
; Publication No. US20030104404A1
; GENERAL INFORMATION:
; APPLICANT: Wise, Carol A.
; TITLE OF INVENTION: Genetic Markers for Autoimmune Disorder
; FILE REFERENCE: TEX871/4-006US/36000
; CURRENT APPLICATION NUMBER: US/10/067,076
; PRIOR FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: 60/287,893
; PRIOR FILING DATE: 2001-05-01
; PRIOR APPLICATION NUMBER: 09/710,693
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 22
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Human Amino Acid
US-10-067-076-22
```

```
Query Match 88.7%; Score 1909.5; DB 15; Length 416;
Best Local Similarity 87.7%; Pred. No. 1.1e-145;
Matches 365; Conservative 20; Mismatches 30; Indels 1; Gaps 1;

QY 1 MMAQLOFRDAFWCRDFTAHGTGYEVLLQRLLDGRRKCKDVEELLRQRAQAEERYGKELVQI 60
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MMPOLQFKDAFWCRDFTAHGTGYEVLLQRLLDGRRKCKDVEELLRQRAQAEERYGKELVQI 60
QY 61 ARKAGGQTEINSLRASFDSLKQOETVNGSAHIQALALREBELSLLEEFREKQEKRYE 120
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 ARKAGGQTEINSLRASFDSLKQOETVNGSSHQIALALREBELSLLEEFREKQEKRYE 120
QY 121 AINDRVQKSKLSLYKTTMESKKAYDQKCRDADDAEQAFERVSANGHOKQVEKSNARQOC 180
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 AINDRVQKSKLSLYKTTMESKKAYDQKCRDADDAEQAFERISANGHOKQVEKSNARQOC 180
QY 181 KESATAEAEVRYRONITOLEARTEMOEHRRTCEAFOLQOEFDRLTLIRNALVHNCQLSM 240
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 KDSATAEAEVRYROSIAOLEKVRAMEOEHRRTCEAFOLQOEFDRLTLIRNALVHNSQLSM 240
QY 241 QCVKDELYEVRITLLEGGDVEGDIINGFIQSKSTGPREPAVPYQNYVDREVPLTSSPG 300
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 241 QCVKDELYEVRITLLEGGDVEGDIINGFIQAKSTGTEPPAPVPYQNYVDREVPLTSSPG 300
QY 301 IOPSCGVIRKRFSGLLHSGSKTTP-SAPAASTETLTPPRNELVYASIEVQATOGNLNS 359
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 301 IOPSCGMIRKRFSGLLHSGSKTTP-SAPAASTETLTPPRNEGYVTAIAVQELQGNPASP 360
QY 360 AODYRALDYTAONSDLDISAGDILAVILLEGEDGMWTVERNQGRGFVGSYLEKL 415
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 361 AOEYRALDYTAONPDELDSLADGILLEVILLEGEDGMWTVERNQGRGFVGSYLEKL 416
```

```
RESULT 4
US-10-067-076-2
; Sequence 2, Application US/10067076
; Publication No. US20030104404A1
; GENERAL INFORMATION:
; APPLICANT: Wise, Carol A.
; TITLE OF INVENTION: Genetic Markers for Autoimmune Disorder
; FILE REFERENCE: TEX871/4-006US/36000
; CURRENT APPLICATION NUMBER: US/10/067,076
; PRIOR FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: 60/287,893
; PRIOR FILING DATE: 2001-05-01
; PRIOR APPLICATION NUMBER: 09/710,693
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 323
; TYPE: PRT
```

```
; ORGANISM: Murine Amino Acid
US-10-067-076-2

Query Match      77.4%; Score 1665; DB 15; Length 323;
Best Local Similarity 99.4%; Pred. No. 3.8e-126;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MAALQFRAFCWCRDFTAGTGYEVLQRLDGRKMKDVEELLRQRAQAEERYGKELVQI 60
   |||||
Db 1 MAALQFRAFCWCRDFTAGTGYEVLQRLDGRKMKDVEELLRQRAQAEERYGKELVQI 60

QY 61 ARRAGQOTEMNSLRTSFDSLKQOTENVGSAHIQALALREELSLSEEFREKQKRYE 120
   |||||
Db 61 ARRAGQOTEMNSLRTSFDSLKQOTENVGSAHIQALALREELSLSEEFREKQKRYE 120

QY 121 AIDRVQKSLSLYKKTMEKRAYDQCRADDAEQAEFVVSANGHOKQVEKSNKAKQC 180
   |||||
Db 121 AIDRVQKSLSLYKKTMEKRAYDQCRADDAEQAEFVVSANGHOKQVEKSNKAKQC 180

QY 181 KESATFAERYRONIQLERARTEMEOEHRTTCAPOLQEPDLTLIRNALWVHCNQLSM 240
   |||||
Db 181 KESATFAERYRONIQLERARTEMEOEHRTTCAPOLQEPDLTLIRNALWVHCNQLSM 240

QY 241 QCVKDELYEEVRLTEGCDVEGDINGFIQSKSTGREPPAPVPYQNYVDEEVTPILGISPS 300
   |||||
Db 241 QCVKDELYEEVRLTEGCDVEGDINGFIQSKSTGREPPAPVPYQNYVDEEVTPILGISPS 300

QY 301 IOPSCGVIRKFSGLLHSGPKTTP 323
   |||||
Db 301 IOPSCGVIRKFSGLLHSGPKTTP 323

RESULT 5
US-10-067-076-3
; Sequence 3, Application US/10067076
; Publication No. US20030104404A1
; GENERAL INFORMATION:
; APPLICANT: Mize, Carol A.
; TITLE OF INVENTION: Genetic Markers for Autoimmune Disorder
; FILE REFERENCE: TEX871/4-006US/36000
; CURRENT FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: 60/287,893
; PRIOR FILING DATE: 2001-05-01
; PRIOR APPLICATION NUMBER: 09/710,693
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Murine Amino Acid
US-10-067-076-3

Query Match      22.0%; Score 473; DB 15; Length 92;
Best Local Similarity 100.0%; Pred. No. 6.8e-31;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 SAPAASTETLTPPERNELVYASIEVOATGNNSSAQDYRALYDTAONSDELIDISACD 383
   |||||
Db 1 SAPAASTETLTPPERNELVYASIEVOATGNNSSAQDYRALYDTAONSDELIDISACD 60

QY 384 ILAVILEGEGDWMTVERNGRGFVPGSYLEKL 415
   |||||
Db 61 ILAVILEGEGDWMTVERNGRGFVPGSYLEKL 92

RESULT 6
US-09-764-868-1021
; Sequence 1021, Application US/09764868
; Patent No. US2002016871A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
```

```
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT32
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1021
; LENGTH: 209
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (148)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (150)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (153)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (154)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
US-09-764-868-1021

Query Match      19.3%; Score 415.5; DB 10; Length 209;
Best Local Similarity 48.3%; Pred. No. 9.1e-26;
Matches 83; Conservative 33; Mismatches 55; Indels 1; Gaps 1;

QY 2 MAALQFRAFCWCRDFTAGTGYEVLQRLDGRKMKDVEELLRQRAQAEERYGKELVQI 61
   |||||
Db 27 MTRSLFRGNFMSADILSTIGDNIQHLNNRKNCKEEDLKEKAAIEERYGDLNLS 86

QY 62 RKAG-GOTEMNSLRTSFDSLKQOTENVGSAHIQALALREELSLSEEFREKQKRYE 120
   |||||
Db 87 RKPCGQSEINTLTKRALEVPKQVDNVAQCHIQLAQSIREBARKEEFREKQKRYE 146

QY 121 AIDRVQKSLSLYKKTMEKRAYDQCRADDAEQAEFVVSANGHOKQVEK 172
   |||||
Db 147 LKMAIYXKSLQFRKTTWDARKNEYQKCRDDEAQSANLVNPKQOKR 198

RESULT 7
US-10-198-070-60
; Sequence 60, Application US/10198070
; Publication No. US20030109437A1
; GENERAL INFORMATION:
; APPLICANT: AVERBACK, PAUL
; TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER
; TITLE OF INVENTION: CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF
; FILE REFERENCE: 59003, 000008
; CURRENT FILING DATE: US/10/198,070
; CURRENT APPLICATION NUMBER: 60/306,150
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: 60/306,161
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: 60/331,477
; PRIOR FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 125
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 60
; LENGTH: 177
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-198-070-60

Query Match      14.2%; Score 305.5; DB 15; Length 177;
Best Local Similarity 48.8%; Pred. No. 5.2e-17;
Matches 59; Conservative 26; Mismatches 35; Indels 1; Gaps 1;
```


RESULT 10

US-10-025-380-685

Sequence 685, Application US/10025380
Publication No. US20020182191A1

GENERAL INFORMATION:

APPLICANT: Xu, Jiangchun

APPLICANT: Lodes, Michael J.

APPLICANT: Secrist, Heather

APPLICANT: Benson, Darin R.

APPLICANT: Meagher, Madeleine Joy

APPLICANT: Stolk, John A.

APPLICANT: Wang, Tongtong

APPLICANT: Smith, Carole L.

APPLICANT: King, Gordon E.

APPLICANT: Wang, Aljun

APPLICANT: Clapper, Jonathan D.

APPLICANT: Skeiky, Yasir A. W.

APPLICANT: Fanger, Gary R.

APPLICANT: Vedrick, Thomas S.

APPLICANT: Carter, Darick

TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS

TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE

FILE REFERENCE: 210121.471C14

CURRENT FILING DATE: 2001-12-19

CURRENT FILING DATE: 2001-12-19

NUMBER OF SEQ ID NOS: 1129

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 685

LENGTH: 486

TYPE: PRM

ORGANISM: Homo sapiens

US-10-025-380-685

Query Match 13.2%; Score 284.5; DB 14; Length 486;

Best Local Similarity 21.1%; Pred. No. 1.1e-14; Mismatches 102; Conservative 92; Mismatches 200; Indels 89; Gaps 14;

Matches 102; Conservative 92; Mismatches 200; Indels 89; Gaps 14;

9 DATECRDFTMHTG-YEVLLORLIDGRMKCKDVEELLROAQAERYGKELVOIARKAGQ 67

16 DSEW-----EVCNRYKRTVKRIDGHRILCSIDLMLCHERRARIEKAYAQOLEMARWROL 69

68 TE-----NLSLRTSDLSKQOTENYNSAHIOALALRELRLEFEHREKQK----- 114

70 VEGPOYGTVEKAMAMMSEAEVSELEHLEVKASLMD--DPEKIKMOKAEFRKOMMG 127

115 --ORRKYEAIMDRYOKSKLSLYKKTMEKSKAYDQKCRDADA--EQAPEFVSANGHOKOV 170

128 FKREKEDDFRKAQOKMAKKIEVEAKKAHNAACKKEKLAISREANSKADPSNPOL 187

171 EKSONKAKOCKESATEAEARYRONIEOLERARTWEGEHRTTCEAFOLQDFRILTLRNA 230

188 KKLQDKIEKCKQDVLTKREKYEKSLKELDQSTPYEMNEQVFECCQFEKRLFEFFREV 247

231 LWHNCLNSQCYVD-DELYEEVRLTLEGCEVEDINGFIOKSTGRRPPAPVPIQNTYD 289

248 LLEVOKHLNLSNAGYKAIYHDELOSIRAAVADELRFRANHGGMAMNP-----QFE 302

290 REVYPLIGSPSIOPS-----CGVIKRPFGILHGSPTT---PSAPAST----- 330

303 EMSADLRLTSLRREKKAATGFTLTGINTGDOFLPSKPSSTLVNPSNPASOASSYN 362

331 -----ETLAPPERNELVYASIP-----VOATQGLNNSAOD-- 362

363 PFEDDEDGTGTSVEKEKIKAKNWSYKTSYPTDWSDESNNPFSSTDANGDSNPFDD 422

363 -----YALVDYTAQNSDELISAGDILAVILEGDDGWYTER--NGORGFPVPSYL 412

423 ATSGTEVRYRALVDYEGOEHELSFKAGDELTKMEDDEGCMCKGRLLDNGOVGLYPAWYV 482

413 EKL 415

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB 483 EAI 485

RESULT 11

US-09-867-550-1862

Sequence 1862, Application US/09867550

Patent No. US20020082206A1

GENERAL INFORMATION:

APPLICANT: Leach, Martin D.

APPLICANT: Mehraban, Fuad,

APPLICANT: Conley, Pamela

APPLICANT: Law, Debbie

APPLICANT: Topper, James

TITLE OF INVENTION: No. US20020082206A1el Polynucleotides from Atherogenic Cells a

TITLE OF INVENTION: Thereby

FILE REFERENCE: 21402-013 (Cura-313)

CURRENT APPLICATION NUMBER: US/09/867,550

CURRENT FILING DATE: 2001-09-20

PRIOR APPLICATION NUMBER: USSN 60/208,427

PRIOR FILING DATE: 2000-05-30

NUMBER OF SEQ ID NOS: 2125

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 1862

LENGTH: 122

TYPE: PRM

ORGANISM: Homo sapiens

US-09-867-550-1862

Query Match 13.0%; Score 280; DB 9; Length 122;

Best Local Similarity 45.9%; Pred. No. 3.5e-15; Mismatches 56; Conservative 21; Mismatches 45; Indels 0; Gaps 0;

Matches 56; Conservative 21; Mismatches 45; Indels 0; Gaps 0;

107 EFEROKQKRYEAIMDRYOKSKLSLYKKTMEKSKAYDQKCRDADAQAPEFVSANGH 166

1 EFEROKQKQKRTKTEIMDAIHOKSLOPKFTMDAKKNYQKCRDQEAQAVSRANLVN 60

167 QKQVKSQKAKOCKESATEAEARYRONIEOLERARTWEGEHRTTCEAFOLQDFRILTL 226

61 PQOERLVPKLTASTVAEDSDKAYMLHIGTLDKVREWQSEHITKACEAFQACERINF 120

227 LR 228

121 FR 122

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

US-09-879-957-34

Sequence 34, Application US/09879957

Patent No. US20020034755A1

GENERAL INFORMATION:

APPLICANT: SPARKS, Andrew B.

HOFFMAN, NO. US20020034755A1h

KAY, Brian K.

FOLKES, Dana M.

MCCONNELL, Stephen J.

TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL

DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND

USING SAME

NUMBER OF SEQUENCES: 227

CORRESPONDENCE ADDRESS:

ADDRESS: Pennie & Edmonds LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/879,957
FILING DATE: 13-Jun-2001
CLASSIFICATION: <Unknown>
PRIORITY INFORMATION:
APPLICATION NUMBER: US 08/630,915
FILING DATE: 03-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 441 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-09-879-957-34

Query Match 12.3%; Score 265.5; DB 9; Length 441;
Best Local Similarity 21.2%; Pred. No. 3.1e-13;
Matches 100; Conservative 84; Mismatches 192; Indels 95; Gaps 16;

3 AQLQFDFWCFRCFTHTG-YEVLLQRLDGRKCKDVEELLQKQAEERYKELVQIA 61
8 ASSEITDSFV-----EVGNKRTYKRIDGHRCLNDLMSCVORAKIEKAYAOQLDWA 61
62 RK-----AGGQTEMNLSRTSPSLKQOTENVSANLQIALALREELSLSEEREOKE- 114
62 KRWRLIEKGPQ--YGSLEFAMGAMTLEADKVSLEHGVKNSLNE--DLEKYNMOKDA 117
115 -----QKKYALMDRVQKSKLSLYKTMESKAYDOKCRADDA--EQAERYS 162
118 YHKQINGFKETKEADGFRKAKQPKAKMKLEAKKAYHLACKERLMTREMSKTE 177
163 ANGHOKVEKSONKAKCKESATEAERYRONIEOLERARTEOEHRRTCEAFQOEED 222
178 GSVTEPQOKKLVKDYKCKQDYOKTEKEKYLEVDYKTPOTMEGMEOVFFQCCQFEK 237
223 RLTTLR-----NALMVHQNLSMQCVKDELYEVRLLTEGCDVEGDING 268
238 RLVLKEVLLDIRHLNLAENSSYMH-----VYRELEGAIRGAAQEDLRNF 284
269 IOSKSTGREPPAPVYQNYDYREVPLIGSPSIQSPGVIKRFSGLLHGSPTTPSAPAA 328
285 RSTSGPGM--PNNWPOFEEMNDLPHTTAKKREKOP--KRAEG-----ATLSNATG 330
339 STELLTPPERNEL-----VYASIEVQATQGN-----LNSAODY-----RAL 366
331 AVESTSQAQDGSVSSYDRGQTYATENWSDDESGNFGFGANGANGANPEEDAKGVYRVL 390
367 YDYTAQNDDELDISAGDILAVILBEGDGMWTVYR--NGORGFVPSYLEKL 415
391 YDYGQEDDELSPKAGDELTKLIGEDDEQGCGRGLDSQLGLYPANYEAI 441

RESULT 13

US-09-879-957-36
Sequence 36, Application US/09879957
Patent No. US20020034755A1
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
HOFFMAN, No. US20020034755A1h
KAY, Brian K.
FOWLES, Dana M.
MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND

USING SAME
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/879,957
FILING DATE: 13-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/630,915
FILING DATE: 03-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 377 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-09-879-957-36

Query Match 9.2%; Score 197; DB 9; Length 377;
Best Local Similarity 20.7%; Pred. No. 8.2e-08;
Matches 79; Conservative 64; Mismatches 165; Indels 74; Gaps 10;

91 HTQLALREELSLSEEREOKEQKRYEALMDRVQKSKLSLYKTMESKAYDOKRD 150
12 HQMNGFKETKEADGFRKAKQPKAKR-----LKEVPAKKAHHAACK 56
151 ADDA--EQAERYSANGHOKVEKSONKAKCKESATEAERYRONIEOLERARTEOE 208
57 EKLATSRANSKADSLNPEOLKRLQDKTEKCKQDYLTKEKYSLSKELDQGPOTYEN 116
209 HRTTEAFQLOEFDRLLTRNALMVHQNLSMQCVK--DELYEVRLLTEGCDVEGDING 267
117 MEQVEEOOQOEKRLFRREVLLLEVOKHLDSNVAGYKALYHDEQSIIRAADVEDLR 176
268 FIOSKSTGREPPAPVYQNYDYREVPLIGSPSIQSPGVIKRFSGLLHGSPTTP 321
177 FRANGPGM--AMNWPQFEEMNDLNTLSRREKKSSTDGVTLTGINTQDOSLPSPSS 234
322 T--PSAPASTETLT-----PTPERNELVYASIE----- 348
235 TLNVPNSPQAQSSVNPFEDEDGTGYSKEDDTKAKKAVSYEXTQSTPTWSDDES 294
349 ---VQATQGNLNSAOD-----YRALYDYTAQNSDELDISAGDILAVILBEGDGM 395
295 NNPFSSSTDANDSNPFDDATSGTEVVRVRLALYDEQGEHDELSPKAGDELTKMDEDEQ 354
396 WTVYR--NGORGFVPSYLEKL 415
355 WCKGRDLNGOGLYPANYEAI 376

RESULT 14

```

US-09-879-957-30
; Sequence 30, Application US/09879957
; Patent No. US2002003475A1
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
;   HOFFMAN, No. US20020034755A1h
;   KAY, Brian K.
;   FOMUKES, Dana M.
;   MCCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
;   DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
;   USING SAME
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Pennie & Edmonds LLP
;   STREET: 1155 Avenue of the Americas
;   CITY: New York
;   STATE: New York
;   COUNTRY: USA
;   ZIP: 10036-2711
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/879,957
;   FILING DATE: 13-Jun-2001
;   CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/630,915
;     FILING DATE: 03-APR-1996
; ATTORNEY/AGENT INFORMATION:
;   NAME: Mistock, S. Leslie
;   REGISTRATION NUMBER: 18,872
;   REFERENCE/DOCKET NUMBER: 1101-174
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 790-9090
;   TELEFAX: (212) 869-8864/9741
;   TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 30:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 788 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: <Unknown>
;     TOPOLOGY: unknown
;     MOLECULE TYPE: peptide
;     SEQUENCE DESCRIPTION: SEQ ID NO: 30:
US-09-879-957-30
Query Match      8.4%   Score 181;   DB 9;   Length 788;
Best Local Similarity 22.5%; Pred. No. 4,5e-06;
Matches 99; Conservative 56; Mismatches 123; Indels 162; Gaps 21
QY      EVLLQRLLDGKRMCDVEELLQRQRAE-----EYKGEIYQI--- 60
DB      462 EFIRRHDDKREKLADQRRLRKREDEADIAARRHTGVIPTHQFTNNEFG-DLNIIDPT 520
QY      61 -ARRAGGOTENNSLRTSPDSLKQOENVGSAHITOLALREPLRSLEEFREPKOKORRY 119
DB      521 AKRSG--LEKRPAPAKFD-EKAQTLK-----ELPLQKGDVY-----YTHQIDQ-NMY 565
QY      120 EAINDRVOKSLSLVKKTM-----ESKRAYDQCKDADAE--QAERYASANGHOKOYER 172
DB      566 EG---EHHGVGIVPRTYIELLPRAEKAQPRKLAIPVQYLEGALAKFNFG-DTQYEM 620
QY      173 SQNAKQCKESATEAEERYVRONIEDLERARTWEDQHRTTCGAFOLQEPDLTLIRNALM 232
DB      621 SFRKG-----EGHITPG-----TSROGIPFTYVDVLR-- 634
QY      233 VHCNQLSMQCYKDDLEYEVRLTLEGCGVEGDINGFIQSKSGREPPAPVQYNYNDREV 292
DB      635 -----DENWY-----EGHITPG-----TSROGIPFTYVDVLR-- 662

```

```

0Y      293 PPLIGSP-----SIOPSGCVIRFSLGLHGSKTPSPAASAEITLPTPENNELV   34
        ||::||| |::||| |::||| |::||| |::||| |::||| |::||| |::|||
Db      663 -PLVETPVDIIDLPRSSSPS-----RSATVSQAASHHSISAGPDLTESK   706
        ||::||| |::||| |::||| |::||| |::||| |::||| |::||| |::|||
0Y      345 ASIEVOATQGNLNSSAQ-----DYRALDYTAONSDLELDISAGDILAVILEGEGWM-   396
        ::||| |::||| |::||| |::||| |::||| |::||| |::||| |::|||
Db      707 NYVGOAQAORRYTDPDRSQPLDLCGYALYSVPQNDELTELRODDIYDVMEKCODGWVF   766
        ||::||| |::||| |::||| |::||| |::||| |::||| |::||| |::|||
0Y      397 -TVERNRGCRGFVPGSYLEKL 415
        |::||| |::||| |::||| |::||| |::||| |::||| |::||| |::|||
Db      767 GTSRRTRQFGTFPPGMNVKPL 786

RESULT 15
US-09-879-957-18
; Sequence 18, Application US/09879957
; Patent No. US20020034755A1
; GENERAL INFORMATION:
    APPLICANT: SPARKS, Andrew B.
                HOFFMAN, No. US20020034755A1h
                RAY, Brian K.
                FOWLER, Dana M.
    TITLE OF INVENTION: MCCONNELL, Stephen J.
                        POLYPEPTIDES HAVING A FUNCTIONAL
                        DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
                        USING SAME
    NUMBER OF SEQUENCES: 227
    CORRESPONDENCE ADDRESSES:
        ADDRESSER: Pennie & Edmonds LLP
        STREET: 1155 Avenue of the Americas
        CITY: New York
        STATE: New York
        COUNTRY: USA
        ZIP: 10036-2711
    COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk
        COMPUTER: IBM PC compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE: PatentIn Release #1.0, Version #1.30
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/09/879,957
        FILING DATE: 13-Jun-2001
        CLASSIFICATION: <Unknown>
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 08/630,915
        FILING DATE: 03-APR-1996
        ATTORNEY/AGENT INFORMATION:
            NAME: Mastrock, S. Leslie
            REGISTRATION NUMBER: 18,872
            REFERENCE/DOCKET NUMBER: 1101-174
        TELECOMMUNICATION INFORMATION:
            TELEPHONE: (212) 790-9090
            TELEFAX: (212) 869-8864/9741
            TELE: 66141 PENNIE
    INFORMATION FOR SEQ ID NO: 18:
        SEQUENCE CHARACTERISTICS:
            LENGTH: 433 amino acids
            TYPE: amino acid
            STRANDEDNESS: <Unknown>
            TOPOLOGY: unknown
        MOLECULE TYPE: peptide
        SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-09-879-957-18

Query Match          8.3%: Score 179.5, DB 9, Length 433;
Best Local Similarity 20.2%; Pred. No. 2,6e-06;
Matches 80; Conservative 59; Mismatches 145; Indels 113; Gaps 14

0Y      47 AOAEEERYKEELV--OIARAKAGQGTEMNSLTSTSFDSLKXQTENVGSAAHQLALALNEELRS 104
        |||||| |::||| |::||| |::||| |::||| |::||| |::||| |::|||
Db      117 ARADEEDVEECIMEWVAAGANNVSHHESTSEFOVGQPAP-VGSGV-----OKTNA 167

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2003, 09:17:57 ; Search time 86 Seconds

(without alignments)
765.948 Million cell updates/sec

Title: US-09-068-377C-1

Perfect score: 415

Sequence: 1 MAAQLQFRDAFWCRDFAHT.....WVERNGRGVPGSYLEKL 415

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

A.GeneSeq_19Jun03:*

- 1: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT:*
- 2: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT:*
- 3: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT:*
- 4: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT:*
- 5: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT:*
- 6: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT:*
- 7: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT:*
- 8: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT:*
- 9: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT:*
- 10: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT:*
- 11: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT:*
- 12: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT:*
- 13: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT:*
- 14: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT:*
- 15: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT:*
- 16: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT:*
- 17: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT:*
- 18: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT:*
- 19: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT:*
- 20: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT:*
- 21: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT:*
- 22: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT:*
- 23: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT:*
- 24: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	415	100.0	415	19	AAW71595
2	415	100.0	415	21	AAB08486
3	415	100.0	415	21	AAV81928
4	31	7.5	397	20	AAV28292
5	31	7.5	400	21	AAV81929
6	31	7.5	416	20	AAV28291
7	31	7.5	416	21	AAV68782
8	10	2.4	4623	22	ABB71106
9	8	1.9	344	21	AAV25760

10	8	1.9	344	21	AAV38295	Arabidopsis thalia
11	8	1.9	344	21	AAV38296	Arabidopsis thalia
12	8	1.9	348	23	AAV30935	Human novel protei
13	8	1.9	348	23	AAE21721	Human FKIN-16 prot
14	8	1.9	425	22	AAV90262	C glutamicum prote
15	8	1.9	440	21	AAV95060	Candida albicans p
16	8	1.9	452	22	AAV62691	Propionibacterium
17	8	1.9	457	22	ABV69700	Drosophila melano
18	8	1.9	474	23	ABP74010	Candida albicans e
19	8	1.9	509	22	AAU02832	Taxus cuspidata ox
20	8	1.9	513	23	ABP65282	Blifidobacterium 10
21	8	1.9	572	22	AAU03506	Human protein kina
22	8	1.9	1139	23	ABU05809	M. tuberculosis an
23	8	1.9	1180	22	ABG17378	Novel human diagno
24	7	1.7	57	17	AAV34234	SH3 domain D of hu
25	7	1.7	59	24	AAV79980	Human P85A residue
26	7	1.7	66	21	AAV44099	Arabidopsis thalia
27	7	1.7	70	22	ABG49386	Human liver peptid
28	7	1.7	70	22	ABR29392	Peptide #2043 enco
29	7	1.7	70	22	ABR34569	Peptide #2075 enco
30	7	1.7	70	22	AAV55356	Human brain expres
31	7	1.7	70	22	AAV15561	Peptide #1995 enco
32	7	1.7	70	22	AAV28051	Peptide #2088 enco
33	7	1.7	70	22	AAV03307	Peptide #1989 enco
34	7	1.7	70	23	ABG37306	Human peptid enco
35	7	1.7	73	23	AAV69580	Cell death protect
36	7	1.7	82	22	AAU40078	Propionibacterium
37	7	1.7	86	22	AAV74950	Bovine P13-SH3 am1
38	7	1.7	90	21	AAV45486	Bovine secreted pro
39	7	1.7	90	21	AAV45486	Human secreted pro
40	7	1.7	111	22	AAU17437	Novel signal trans
41	7	1.7	119	22	AAV58092	Propionibacterium
42	7	1.7	120	21	AAV30027	Arabidopsis thalia
43	7	1.7	123	21	AAV38276	Arabidopsis thalia
44	7	1.7	127	22	ABG04777	Novel human diagno
45	7	1.7	129	21	AAV01561	Human secreted pro

ALIGNMENTS

RESULT 1	AAW71595	standard; protein; 415 AA.
AC	AAW71595;	
XX		
DT	24-NOV-1998 (first entry)	
XX		
DE	Murine tyrosine phosphorylated cleavage furrow-associated protein.	
KW	Mouse: tyrosine phosphorylated cleavage furrow-associated protein;	
KW	PSRTP; PEST family; protein tyrosine phosphatase; murine;	
KW	polymerisation; actin monomer; eukaryotic cell; identification;	
KW	antagonist.	
XX		
OS	Mus sp.	
XX		
PN	W09835037-A1.	
XX		
PD	13-AUG-1998.	
XX		
PF	30-JAN-1998; 98MO-US01774.	
XX		
PR	29-SEP-1997; 97US-0938829.	
PR	07-FEB-1997; 97US-0798419.	
XX		
PA	(GETH) GENENTECH INC.	
XX		
PI	Dowbenko DJ, Lasky LA;	
XX		
XX	WPI; 1998-447234/38.	
DR	N-PSDB; AAV57973.	

XX New PEST-type protein tyrosine phosphatase interacting polypeptide -
PT nucleic acids and vectors; for inducing the polymerisation of actin
PT monomers in eukaryotic cells and identifying antagonists
XX
XX
PS Claim 1; Fig 1; 11pp; English.
XX
CC The present sequence represents murine tyrosine phosphorylated cleavage
CC furrow-associated protein (PSTPIP), which is a PEST-type protein.
CC tyrosine phosphatase (PTP) interacting polypeptide. PSTPIP induces the
CC polymerisation of actin monomers in a eukaryotic cell, by introducing
CC a vector containing the nucleic acid sequence encoding PSTPIP into the
CC cell. Assays for identifying (ant)agonists of PSTPIP comprise contacting
CC PSTPIP with the agent and monitoring the ability of PSTPIP to induce
CC actin polymerisation.
XX
SQ Sequence 415 AA:

Query Match 100.0%; Score 415; DB 19; Length 415;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMAQLQFRDAFWCRDFTAHGTGVEVLLQRLDGRKMKDVEELLRQRAQAEERKGELVQI 60
DB 1 MMAQLQFRDAFWCRDFTAHGTGVEVLLQRLDGRKMKDVEELLRQRAQAEERKGELVQI 60
QY 61 ARKAGGQTEMNSLRTSFDSLKQOTENVGSAHIQALALRELSLEEFREKQEKRYE 120
DB 61 ARKAGGQTEMNSLRTSFDSLKQOTENVGSAHIQALALRELSLEEFREKQEKRYE 120
QY 121 AIMDRVOKSKLSLYKKTMSKRAYDOKCRDADDAEQAFERVSANGHOKOVEKSQNKAKOC 180
DB 121 AIMDRVOKSKLSLYKKTMSKRAYDOKCRDADDAEQAFERVSANGHOKOVEKSQNKAKOC 180
QY 181 KESATEAEVRYRONITOLERARTEMOEHRITCEAFOLOEFDDLTLRNALWCHQOLSM 240
DB 181 KESATEAEVRYRONITOLERARTEMOEHRITCEAFOLOEFDDLTLRNALWCHQOLSM 240
QY 241 QCVKDELEYEVRLLTEGCDVEGDINGFIQSKSTGREPAPYQNYVDREVTPILGSPS 300
DB 241 QCVKDELEYEVRLLTEGCDVEGDINGFIQSKSTGREPAPYQNYVDREVTPILGSPS 300
QY 301 IOPSCGVIRKFSGLLHGSPKTPPSAPAASTETLTPPERNELVYASIEVOATOGNINSSA 360
DB 301 IOPSCGVIRKFSGLLHGSPKTPPSAPAASTETLTPPERNELVYASIEVOATOGNINSSA 360
QY 361 ODYRALYDTAONSDELISAGDILAVILLEGGDGMWTVERNGORGFPVPSYLEKL 415
DB 361 ODYRALYDTAONSDELISAGDILAVILLEGGDGMWTVERNGORGFPVPSYLEKL 415

RESULT 2
AAB08486 standard; Protein; 415 AA.
XX
AC AAB08486;
XX
DT 20-DEC-2000 (first entry)
XX
DE A murine PTP phosphatase interacting protein (PSTPIP).
XX
KM protein tyrosine phosphatase; PTP; phosphatase interacting protein;
KM PSTPIP; PEST family; protein tyrosine phosphatase; actin monomer;
KM tissue typing; tumour cell; tumour imaging.
XX
OS Mus sp.
XX
PN US611073-A.
XX
PD 29-AUG-2000.
XX
PF 06-FEB-1998; 98US-0020222.
XX

PR 07-FEB-1997; 97US-0104590.
XX
XX (GENTH) GENENTECH INC.
XX
XX Lasky LA;
XX
XX WPI; 2000-586378/55.
XX
XX DR N-PSDB; AAA64240.
XX
PT Novel PST phosphatase interacting protein useful for inducing
PT polymerisation of actin monomers and for identifying homologue of PST
PT phosphatase interacting protein
XX
XX Claim 1; Column 45-48; 48pp; English.
XX
CC The present sequence represents a protein tyrosine phosphatase (PTP)
CC phosphatase interacting protein (PSTPIP). PSTPIP polypeptides are
CC bound by and dephosphorylated by the PEST family of protein tyrosine
CC phosphatases. PSTPIP associates with actin. PSTPIP is useful for
CC inducing the polymerisation of actin monomer in eukaryotic cells by
CC introducing the polypeptide into the cell. The polypeptide is useful
CC for identifying and isolating PSTPIP homologues in another mammalian
CC species. In screening assays to identify antagonists and agonists of
CC native PSTPIP polypeptide and as molecular weight markers on protein
CC gels. The PSTPIP nucleic acid is useful for tissue typing of specific
CC mammalian tissues, for preparing PSTPIP polypeptides by recombinant
CC techniques, as hybridisation probes for searching cDNA and genomic
CC libraries for the coding sequence of other PSTPIP analogues and to
CC isolate homologous genes specifically expressed in tumour cells.
CC Antagonists of PSTPIP peptide are useful for inhibiting biological
CC activity of the peptide. Antibodies of PSTPIP are useful to identify
CC rapidly dividing cells and are used to image tumours comprising such
XX
SQ Sequence 415 AA:

Query Match 100.0%; Score 415; DB 21; Length 415;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMAQLQFRDAFWCRDFTAHGTGVEVLLQRLDGRKMKDVEELLRQRAQAEERKGELVQI 60
DB 1 MMAQLQFRDAFWCRDFTAHGTGVEVLLQRLDGRKMKDVEELLRQRAQAEERKGELVQI 60
QY 61 ARKAGGQTEMNSLRTSFDSLKQOTENVGSAHIQALALRELSLEEFREKQEKRYE 120
DB 61 ARKAGGQTEMNSLRTSFDSLKQOTENVGSAHIQALALRELSLEEFREKQEKRYE 120
QY 121 AIMDRVOKSKLSLYKKTMSKRAYDOKCRDADDAEQAFERVSANGHOKOVEKSQNKAKOC 180
DB 121 AIMDRVOKSKLSLYKKTMSKRAYDOKCRDADDAEQAFERVSANGHOKOVEKSQNKAKOC 180
QY 181 KESATEAEVRYRONITOLERARTEMOEHRITCEAFOLOEFDDLTLRNALWCHQOLSM 240
DB 181 KESATEAEVRYRONITOLERARTEMOEHRITCEAFOLOEFDDLTLRNALWCHQOLSM 240
QY 241 QCVKDELEYEVRLLTEGCDVEGDINGFIQSKSTGREPAPYQNYVDREVTPILGSPS 300
DB 241 QCVKDELEYEVRLLTEGCDVEGDINGFIQSKSTGREPAPYQNYVDREVTPILGSPS 300
QY 301 IOPSCGVIRKFSGLLHGSPKTPPSAPAASTETLTPPERNELVYASIEVOATOGNINSSA 360
DB 301 IOPSCGVIRKFSGLLHGSPKTPPSAPAASTETLTPPERNELVYASIEVOATOGNINSSA 360
QY 361 ODYRALYDTAONSDELISAGDILAVILLEGGDGMWTVERNGORGFPVPSYLEKL 415
DB 361 ODYRALYDTAONSDELISAGDILAVILLEGGDGMWTVERNGORGFPVPSYLEKL 415

RESULT 3
AAY81928 standard; Protein; 415 AA.
XX
XX

AC AAY81928;
XX 27-JUN-2000 (first entry)
XX
XX Murine PST phosphatase interacting protein.
DE
XX PST phosphatase interacting protein; PSTPIP; tumour therapy;
KW protein tyrosine phosphatase; mouse.
XX
XX Mus sp.
OS
XX US6040437-A.
XX
XX 21-MAR-2000.
XX
XX 29-SEP-1997; 97US-0938830.
XX
XX 07-FEB-1997; 97US-0104590.
XX
XX (GETH) GENENTECH INC.
XX
XX Dowdenko DJ, Lasky LA;
XX
XX WPI: 2000-282393/24.
XX
XX N-PSDB; AAA07276.
XX
XX Novel genes encoding protein tyrosine phosphatase binding proteins
PT useful for isolating homologous genes, e.g. in tumour cells, which
PT provide more specific targets for tumour therapy
XX
XX Disclosure; Column 61-64; 65pp; English.
XX
XX This sequence represents the PST phosphatase interacting protein (PSTPIP)
CC sequence of the invention. The protein is a protein tyrosine
CC phosphatase that possesses a non-catalytic domain comprising a proline,
CC serine and threonine rich region and a C-terminal segment of 20 amino
CC acid (aa's) rich in proline, and defines an SH3 binding domain. Nucleic
CC acids encoding native PSTPIP molecules can be used to isolate homologous
CC genes specifically expressed in tumour cells, which might provide more
CC specific targets for tumour therapy. The DNA is also useful for the
CC preparation of PSTPIP polypeptides by recombinant techniques and as
CC hybridisation probes for searching cDNA and genomic libraries for the
CC coding sequence of other PSTPIP polypeptide analogues in other species.
XX
XX Sequence 415 AA:
SQ
Query Match 100.0%; Score 415; DB 21; Length 415;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MAAOQDFDARFMCNDRFTAHGTGEVLLQRLDGRKMKDVEELLRQAAEERYGKEVQI 60
DB 1 MAAOQDFDARFMCNDRFTAHGTGEVLLQRLDGRKMKDVEELLRQAAEERYGKEVQI 60
OY 61 ARKAGGQTEMNSLRTSPDSLKQQTENVSANHIQIALALREELRSLEEFREKQEKRYE 120
DB 61 ARKAGGQTEMNSLRTSPDSLKQQTENVSANHIQIALALREELRSLEEFREKQEKRYE 120
OY 121 AIMRVQSKSLTKYKTKMESKKAIDQKCRDADDAQATERRVSANOHQVEKSQKAKOC 180
DB 121 AIMRVQSKSLTKYKTKMESKKAIDQKCRDADDAQATERRVSANOHQVEKSQKAKOC 180
OY 181 KESATEAERYVRONIEOLERATEMEQEHRTTCEAFQLOEFDRLTLIRNALMVHGNOLSM 240
DB 181 KESATEAERYVRONIEOLERATEMEQEHRTTCEAFQLOEFDRLTLIRNALMVHGNOLSM 240
OY 241 QCVHDELVEEVRLLTEGCDVEGDIINGFIQSKSTGREPPAPVYQYNYDREVTPLIGSPS 300
DB 241 QCVHDELVEEVRLLTEGCDVEGDIINGFIQSKSTGREPPAPVYQYNYDREVTPLIGSPS 300
OY 301 IQPSCGVTKRRSGLLHSGPKTTPSPAPASTETLTPTPRNELVYASIEVQATQGNLNSA 360
DB 301 IQPSCGVTKRRSGLLHSGPKTTPSPAPASTETLTPTPRNELVYASIEVQATQGNLNSA 360

OY 361 ODPRALYDXTAONSDDELISAGDILAVILEGEDGWTVERNGRGVPSYLEKL 415
DB 361 ODPRALYDXTAONSDDELISAGDILAVILEGEDGWTVERNGRGVPSYLEKL 415
RESULT 4
ID AAY28292 standard; protein; 397 AA.
XX
XX AAY28292;
XX
XX 29-SEP-1999 (first entry)
XX
XX Amino acid sequence for CD2BPIs, variant of CD2BPI.
XX
XX Lymphocytes; immune response; CD2; protein; variant;
KW antigen recognition; cell adhesion; amino acid.
XX
XX Homo sapiens.
XX
XX WO936534-A1.
XX
XX 22-JUL-1999.
XX
XX 14-DEC-1998; 98WO-US26699.
XX
XX 13-JAN-1998; 98US-0006428.
XX
XX (DAND) DANA FARBER CANCER INST INC.
XX
XX An W, Li J, Nishizawa K, Reinherz EL;
XX
XX WPI: 1999-444396/37.
XX
XX A human CD2 cytoplasmic tail binding protein, CD2BPI and related
PT polynucleotides, useful in identifying agents which modulate signal
PT transduction, cell adhesion or motility
XX
XX Claim 4; Page 90-91; 106pp; English.
XX
XX This protein is a variant of CD2BPI. CD2BPI is a cd215-1like adapter
CC protein. Anti-CD2BPI antibodies can be used to assay for the presence
CC of CD2BPI in a cell or tissue sample. Modulatory agents identified in
CC the methods are used to inhibit or enhance CD2BPI activity.
CC These agents can be used to modulate signal transduction, cell adhesion
CC or motility. CD2BPI inhibitors are useful for enhancing signal
CC transduction, cell adhesion or motility.
CC CD2BPI enhancers have the opposite effect. Mimics of CD2BPI protein
CC activity can also be used to modulate signal transduction, etc
CC (all claimed).
CC Inhibition of CD2BPI is useful for cancer therapy, to augment the
CC immune response to cancer cells and tumours. Enhancement of CD2BPI is
CC useful in immunocompromised individuals and in immunodeficiency related
CC diseases. Such diseases and conditions, include rheumatoid arthritis,
CC juvenile diabetes, systemic lupus erythematosus, and transplantation, to
CC reduce adhesion of the recipient's T cells with tissues in the
CC auto graft.
XX
XX Sequence 397 AA:
SQ
Query Match 7.5%; Score 31; DB 20; Length 397;
Best Local Similarity 100.0%; Pred. No. 6.7e-21;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 204 EWEQEHRTTCEAFQLOEFDRLTLIRNALMVH 234
DB 204 EWEQEHRTTCEAFQLOEFDRLTLIRNALMVH 234
RESULT 5
AAY81929
ID AAY81929 standard; Protein; 400 AA.

```

XX AC AAY81929;
XX DT 27-JUN-2000 (first entry)
XX DE Human PST phosphatase interacting protein.
XX KW PST phosphatase interacting protein; PSTPIP; tumour therapy;
XX KW protein tyrosine phosphatase; human; ss.
XX OS Homo sapiens.
XX PN US6040437-A.
XX PD 21-MAR-2000.
XX PF 29-SEP-1997; 97US-0938830.
XX PR 07-FEB-1997; 97US-0104590.
XX PA (GETH ) GENENTECH INC.
XX PI Dowbenko DJ, Lasky LA;
XX DR WPI: 2000-282393/24.
XX DR N-PSDB; AAA07277.
XX PT Novel genes encoding protein tyrosine phosphatase binding proteins
XX PT useful for isolating homologous genes, e.g. in tumour cells, which
XX PT provide more specific targets for tumour therapy
XX PS Claim 5; Column 81-84; 65pp; English.
XX CC This sequence represents the PST phosphatase interacting protein (PSTPIP)
XX CC sequence of the invention. The protein is a protein tyrosine
XX CC phosphatase that possesses a non-catalytic domain comprising a proline,
XX CC serine and threonine rich region and a C-terminal segment of 20 amino
XX CC acid (aa's) rich in proline, and defines an SH3 binding domain. Nucleic
XX CC acids encoding native PSTPIP molecules can be used to isolate homologous
XX CC genes specifically expressed in tumour cells, which might provide more
XX CC specific targets for tumour therapy. The DNA is also useful for the
XX CC preparation of PSTPIP polypeptides by recombinant techniques and as
XX CC hybridisation probes for searching cDNA and genomic libraries for the
XX CC coding sequence of other PSTPIP polypeptide analogues in other species.
XX SQ Sequence 400 AA;

Query Match 7.5%; Score 31; DB 21; Length 400;
Best Local Similarity 100.0%; Pred. No. 6.8e-21;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 204 EWEDEHRTTCEAFQLOEFDRITLIRNALMVH 234
DB 223 EWEDEHRTTCEAFQLOEFDRITLIRNALMVH 253

RESULT 6
AAY28291
ID AAY28291 standard; protein; 416 AA.
XX AC AAY28291;
XX DT 29-SEP-1999 (first entry)
XX DE Amino acid sequence for CD2BP1L, variant of CD2BP1.
XX KW Lymphocytes; immune response; CD2; protein; variant;
XX KW antigen recognition; cell adhesion; amino acid.
XX OS Homo sapiens.
XX PN WO9936534-A1.

```

```

PD 22-JUL-1999.
XX 14-DEC-1998; 98WO-US26699.
XX 13-JAN-1998; 98US-0006428.
XX (DAND ) DANA FARBER CANCER INST INC.
XX An W, Li J, Nishizawa K, Reinherz EL;
XX WPI: 1999-444396/37.
XX A human CD2 cytoplasmic tail binding protein, CD2BP1 and related
XX polynucleotides, useful in identifying agents which modulate signal
XX transduction, cell adhesion or motility
XX Claim 3; Page 88-89; 106pp; English.
XX This protein is a variant of CD2BP1. CD2BP1 is a cd615-1-like adapter
XX protein. Anti-CD2BP1 antibodies can be used to assay for the presence
XX of CD2BP1 in a cell or tissue sample. Modulatory agents identified in
XX the methods are used to inhibit or enhance CD2BP1 activity.
XX These agents can be used to modulate signal transduction, cell adhesion
XX or motility. CD2BP1 inhibitors are useful for enhancing signal
XX transduction, cell adhesion or motility.
XX CD2BP1 enhancers have the opposite effect. Mimics of CD2BP1 protein
XX activity can also be used to modulate signal transduction, etc
XX (all claimed).
XX Inhibition of CD2BP1 is useful for cancer therapy, to augment the
XX immune response to cancer cells and tumours. Enhancement of CD2BP1 is
XX useful in immunocompromised individuals and in immunodeficiency related
XX diseases. Such diseases and conditions, include rheumatoid arthritis,
XX juvenile diabetes, systemic lupus erythematosus, and transplantation, to
XX reduce adhesion of the recipient's T cells with tissues in the
XX auto graft.
XX SQ Sequence 416 AA;

Query Match 7.5%; Score 31; DB 20; Length 416;
Best Local Similarity 100.0%; Pred. No. 7e-21;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 204 EWEDEHRTTCEAFQLOEFDRITLIRNALMVH 234
DB 204 EWEDEHRTTCEAFQLOEFDRITLIRNALMVH 234

RESULT 7
AAY68782
ID AAY68782 standard; protein; 416 AA.
XX AC AAY68782;
XX DT 16-MAY-2000 (first entry)
XX DE Amino acid sequence of a human phosphorylation effector PHSP-14.
XX KW Human; phosphorylation effector; PHSP; proliferative disorder;
XX KW immune disorder; neuronal disorder.
XX OS Homo sapiens.
XX Key Location/Qualifiers
XX Modified-site 20
XX Modified-site /note= "potential phosphorylation site"
XX Modified-site 53
XX Modified-site /note= "potential phosphorylation site"
XX Modified-site 72
XX Modified-site /note= "potential phosphorylation site"
XX Modified-site 79
XX Modified-site /note= "potential phosphorylation site"
XX Modified-site 97
XX Modified-site /note= "potential phosphorylation site"

```

Accession	Protein Name	Score	DB	Length	Mismatches	Indels	Gaps
FT	Modified-site	104	"potential phosphorylation site"				
FT	Modified-site	140	/note= "potential phosphorylation site"				
FT	Modified-site	183	/note= "potential phosphorylation site"				
FT	Modified-site	185	/note= "potential phosphorylation site"				
FT	Modified-site	211	/note= "potential phosphorylation site"				
FT	Modified-site	274	/note= "potential phosphorylation site"				
FT	Modified-site	312	/note= "potential phosphorylation site"				
FT	Modified-site	318	/note= "potential phosphorylation site"				
FT	Domain	366..384	/note= "potential phosphorylation site"				
FT	Modified-site	381	/note= "SH3 domain"				
FT	Domain	402..414	/note= "potential phosphorylation site"				
FT	Modified-site	411	/note= "SH3 domain"				
FT	Modified-site		/note= "potential phosphorylation site"				
XX	W0200006728-A2.						
XX	10-FEB-2000.						
XX	28-JUL-1999;	99WO-US17132.					
XX	28-JUL-1998;	98US-0123494.					
PR	14-SEP-1998;	98US-0152814.					
PR	14-OCT-1998;	98US-0173482.					
PR	03-NOV-1998;	98US-0106889.					
PR	19-NOV-1998;	98US-0109093.					
PR	22-DEC-1998;	98US-0113796.					
PR	12-JAN-1999;	99US-0173482.					
PR	12-JAN-1999;	99US-0229005.					
XX	(INCYTE) INCYTE PHARM INC.						
PI	Hillman JL, Lal P, Tang YT, Corley NC, Guegler KJ, Baughn MR;						
PI	Patterson C, Bandman O, Au-Young J, Gorgone GA, Yue H, Azimzal Y;						
PI	Reddy R, Lu DAM, Shih LT;						
XX	WPI; 2000-183125/16.						
DR	N-PSDB; AAZ46151.						
XX	Claim 1; Page 95-96; 142pp; English.						
XX	AAV6879-95 and AAV6879-99 represent human phosphorylation effectors						
CC	(PHSP), designated PHSP1-PHSP1 (the protein sequence for PHSP28 is not						
CC	given in the specification). The sequences were isolated from cDNA						
CC	libraries prepared from various human tissues. The PHSP proteins are						
CC	useful for the diagnosis, treatment and prevention of proliferative						
CC	disorders, immune disorders and neuronal disorders. The PHSP proteins						
CC	form pharmaceutical compositions which useful for treating or preventing						
CC	disorders associated with decreased PHSP expression/activity. PHSP						
CC	antagonists are useful for treating or preventing disorders associated						
XX	with increased PHSP expression/activity.						
XX	Sequence 416 AA;						
XX	Query Match 7.5%; Score 31; DB 21; Length 416;						
XX	Best Local Similarity 100.0%; Pred. NO. 7e-21;						
XX	Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0.						

```

Db      204 EWEQEHRTCEAFOLQEERDLITLIRNALMVH 234

RESULT 8
ABB71106
ID      ABB71106 standard; Protein: 4623 AA.
XX
XX      ABB71106;
AC
XX      26-MAR-2002 (first entry)
DT
XX      Drosophila melanogaster polypeptide SEQ ID NO 40110.
DE
XX      Drosophila: developmental biology; cell signalling; insecticide;
KW      pharmaceutical.
OS      Drosophila melanogaster.
XX      WO200171042-A2.
XX      PN
XX      27-SEP-2001.
PD
XX
XX      23-MAR-2001; 2001WO-US09231.
PF
XX      23-MAR-2000; 2000US-191637P.
PR      11-JUL-2000; 2000US-0614150.
XX
XX      (PEKE ) PE CORP NY.
PA
XX      Venter JC, Adams M, Li PWD, Myers EW;
PL      WPI; 2001-656860/75.
DR      N-PSDB; ABL15209.
XX
XX      New isolated nucleic acid detection reagent for detecting 1000 or more
PT      genes from Drosophila and for elucidating cell signalling and cell-cell
PT      interactions -
XX
XX      Disclosure; SEQ ID NO 40110; 21pp + Sequence listing; English.
PS
XX
CC      The invention relates to an isolated nucleic acid detection reagent
CC      capable of detecting 1000 or more genes from Drosophila. The invention is
CC      useful in developmental biology and in elucidating cell signalling and
CC      cell-cell interactions in higher eukaryotes for the development of
CC      insecticides, therapeutics and pharmaceutical drugs. The invention
CC      discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC      sequences (ABB57737-ABB72072).
CC      (ABB57737-ABB72072)
CC      The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pct_sequences.
CC
SQ      Sequence 4623 AA;

Query Match          2.4%; Score 10; DB 22; Length 4623;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      98 LREELRSLEE 107
        |||||
        LREELRSLEE 3391

Db      3382 LREELRSLEE 3391

RESULT 9
AAG25760
ID      AAG25760 standard; Protein: 344 AA.
XX
XX      AAG25760;
AC
XX      17-OCT-2000 (first entry)
DT
XX
DE      Arabidopsis thaliana protein fragment SEQ ID NO: 29955.
XX

```

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
OS Arabidopsis thaliana.
XX EPI033405-A2.
PN
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 990S-0121825.
PR 05-MAR-1999; 990S-0123180.
PR 09-MAR-1999; 990S-0123548.
PR 23-MAR-1999; 990S-0125788.
PR 25-MAR-1999; 990S-0126264.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 08-APR-1999; 990S-0128714.
PR 16-APR-1999; 990S-0129845.
PR 19-APR-1999; 990S-0130077.
PR 21-APR-1999; 990S-0130449.
PR 23-APR-1999; 990S-0130510.
PR 28-APR-1999; 990S-0130891.
PR 30-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 04-MAY-1999; 990S-0132407.
PR 05-MAY-1999; 990S-0132484.
PR 05-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 07-MAY-1999; 990S-0132487.
PR 07-MAY-1999; 990S-0132863.
PR 11-MAY-1999; 990S-0134256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134221.
PR 14-MAY-1999; 990S-0134370.
PR 18-MAY-1999; 990S-0134768.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135353.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.
PR 27-MAY-1999; 990S-0136392.
PR 28-MAY-1999; 990S-0136782.
PR 01-JUN-1999; 990S-0137222.
PR 03-JUN-1999; 990S-0137528.
PR 04-JUN-1999; 990S-0137502.
PR 07-JUN-1999; 990S-0137724.
PR 08-JUN-1999; 990S-0138094.
PR 10-JUN-1999; 990S-0138540.
PR 10-JUN-1999; 990S-0138847.
PR 14-JUN-1999; 990S-0139119.
PR 16-JUN-1999; 990S-0139452.
PR 16-JUN-1999; 990S-0139453.
PR 17-JUN-1999; 990S-0139492.
PR 18-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139455.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139462.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139763.
PR 21-JUN-1999; 990S-0139817.
PR 22-JUN-1999; 990S-0139899.
PR 23-JUN-1999; 990S-0140353.

PR 23-JUN-1999; 990S-0140354.
PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141287.
PR 01-JUL-1999; 990S-0141642.
PR 02-JUL-1999; 990S-0142154.
PR 06-JUL-1999; 990S-0142055.
PR 08-JUL-1999; 990S-0142830.
PR 09-JUL-1999; 990S-0142820.
PR 12-JUL-1999; 990S-0142977.
PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143624.
PR 15-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 20-JUL-1999; 990S-0144684.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 22-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145085.
PR 22-JUL-1999; 990S-0145089.
PR 22-JUL-1999; 990S-0145089.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145218.
PR 23-JUL-1999; 990S-0145224.
PR 26-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145913.
PR 27-JUL-1999; 990S-0145918.
PR 27-JUL-1999; 990S-0145919.
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 02-AUG-1999; 990S-0146389.
PR 03-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 04-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 05-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 06-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147493.
PR 09-AUG-1999; 990S-0147935.
PR 10-AUG-1999; 990S-0148171.
PR 11-AUG-1999; 990S-0148319.
PR 12-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 13-AUG-1999; 990S-0148684.
PR 16-AUG-1999; 990S-0149368.
PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149723.
PR 20-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0149930.
PR 25-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151303.
PR 31-AUG-1999; 990S-0151438.

PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.9%; Score 8; DB 21; Length 344;
Best Local Similarity 100.0%; Pred. No. 61;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 182 ESATEAER 189
 |||||
DB 117 ESATEAER 124

RESULT 10

AAG38295
ID AAG38295 standard; Protein; 344 AA.

XX AAG38295;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 47223.

XX protein identification; signal transduction pathway; metabolic pathway;
KW hydridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.
PD 25-FEB-2000; 2000EP-0301439.
XX PF 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142390.
PR 08-JUL-1999; 990S-0142803.
PR 09-JUL-1999; 990S-0142920.
PR 12-JUL-1999; 990S-0142977.
PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143624.
PR 15-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 20-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 21-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145085.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145089.
PR 22-JUL-1999; 990S-0145192.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145218.
PR 23-JUL-1999; 990S-0145224.
PR 26-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145913.
PR 27-JUL-1999; 990S-0145918.
PR 27-JUL-1999; 990S-0145919.
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 02-AUG-1999; 990S-0146389.
PR 03-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 04-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 05-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 06-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147493.
PR 09-AUG-1999; 990S-0147935.
PR 10-AUG-1999; 990S-0148171.
PR 11-AUG-1999; 990S-0148319.
PR 12-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 13-AUG-1999; 990S-0148684.
PR 16-AUG-1999; 990S-0149368.
PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149723.
PR 20-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0149930.
PR 25-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151303.
PR 31-AUG-1999; 990S-0151438.
PR 01-SEP-1999; 990S-0151930.
PR 07-SEP-1999; 990S-0152363.
PR 10-SEP-1999; 990S-0153070.
PR 13-SEP-1999; 990S-0153758.
PR 15-SEP-1999; 990S-0154018.
PR 16-SEP-1999; 990S-0154039.
PR 20-SEP-1999; 990S-0154779.

PR 22-SEP-1999; 990S-0155139.
PR 23-SEP-1999; 990S-0155486.
PR 24-SEP-1999; 990S-0155569.
PR 28-SEP-1999; 990S-0156458.
PR 29-SEP-1999; 990S-0156596.
PR 04-OCT-1999; 990S-0157117.
PR 05-OCT-1999; 990S-0157753.
PR 06-OCT-1999; 990S-0157865.
PR 07-OCT-1999; 990S-0158029.
PR 08-OCT-1999; 990S-0158232.
PR 12-OCT-1999; 990S-0158369.
PR 13-OCT-1999; 990S-0159293.
PR 13-OCT-1999; 990S-0159294.
PR 13-OCT-1999; 990S-0159325.
PR 14-OCT-1999; 990S-0159330.
PR 14-OCT-1999; 990S-0159331.
PR 14-OCT-1999; 990S-0159337.
PR 14-OCT-1999; 990S-0159638.
PR 18-OCT-1999; 990S-0159584.
PR 21-OCT-1999; 990S-0160741.
PR 21-OCT-1999; 990S-0160767.
PR 21-OCT-1999; 990S-0160768.
PR 21-OCT-1999; 990S-0160770.
PR 21-OCT-1999; 990S-0160814.
PR 21-OCT-1999; 990S-0160815.
PR 22-OCT-1999; 990S-0160980.
PR 22-OCT-1999; 990S-0160981.
PR 22-OCT-1999; 990S-0160989.
PR 25-OCT-1999; 990S-0161404.
PR 25-OCT-1999; 990S-0161405.
PR 25-OCT-1999; 990S-0161406.
PR 26-OCT-1999; 990S-0161359.
PR 26-OCT-1999; 990S-0161360.
PR 26-OCT-1999; 990S-0161361.
PR 28-OCT-1999; 990S-0161920.
PR 28-OCT-1999; 990S-0161992.
PR 28-OCT-1999; 990S-0161993.
PR 29-OCT-1999; 990S-0162142.

Query Match Best Local Similarity 1.9%; Score 8; DB 21; Length 344;

Matches 8; Conservativity 100.0%; Pred. No. 61;

Mismatches 0; Indels 0; Gaps 0;

OY 182 ESATEAER 189
DB 117 ESATEAER 124

RESULT 11

ID AAG38296 standard; Protein; 344 AA.

AC AAG38296;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 47225.

KW protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 990S-0121825.

PR 05-MAR-1999; 990S-0123180.

PR 09-MAR-1999; 990S-0123548.
PR 23-MAR-1999; 990S-0125788-
PR 25-MAR-1999; 990S-0126264.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 08-APR-1999; 990S-0128714.
PR 16-APR-1999; 990S-0129845.
PR 19-APR-1999; 990S-0130077.
PR 21-APR-1999; 990S-0130449.
PR 23-APR-1999; 990S-0130510.
PR 23-APR-1999; 990S-0130891.
PR 28-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 30-APR-1999; 990S-0132407.
PR 04-MAY-1999; 990S-0132484.
PR 05-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 06-MAY-1999; 990S-0132487.
PR 07-MAY-1999; 990S-0132863.
PR 11-MAY-1999; 990S-0134256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134221.
PR 14-MAY-1999; 990S-0134370.
PR 18-MAY-1999; 990S-0134768.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135353.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.
PR 27-MAY-1999; 990S-0136392.
PR 28-MAY-1999; 990S-0136782.
PR 01-JUN-1999; 990S-0137222.
PR 03-JUN-1999; 990S-0137528.
PR 04-JUN-1999; 990S-0137502.
PR 07-JUN-1999; 990S-0137724.
PR 08-JUN-1999; 990S-0138094.
PR 10-JUN-1999; 990S-0138540.
PR 10-JUN-1999; 990S-0138847.
PR 14-JUN-1999; 990S-0138119.
PR 16-JUN-1999; 990S-0139452.
PR 16-JUN-1999; 990S-0139453.
PR 17-JUN-1999; 990S-0139492.
PR 18-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139455.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139462.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139763.
PR 21-JUN-1999; 990S-0139817.
PR 22-JUN-1999; 990S-0139899.
PR 23-JUN-1999; 990S-0140353.
PR 23-JUN-1999; 990S-0140354.
PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141287.
PR 01-JUL-1999; 990S-0141842.
PR 01-JUL-1999; 990S-0142154.
PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142390.
PR 08-JUL-1999; 990S-0142803.
PR 09-JUL-1999; 990S-0142920.
PR 12-JUL-1999; 990S-0142977.
PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143624.

PR 15-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 20-JUL-1999; 990S-0144884.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 21-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145085.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145089.
PR 22-JUL-1999; 990S-0145192.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145218.
PR 23-JUL-1999; 990S-0145224.
PR 26-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145913.
PR 27-JUL-1999; 990S-0145918.
PR 27-JUL-1999; 990S-0145919.
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 02-AUG-1999; 990S-0146389.
PR 03-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 04-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 05-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 06-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147493.
PR 09-AUG-1999; 990S-0147935.
PR 10-AUG-1999; 990S-0148171.
PR 11-AUG-1999; 990S-0148319.
PR 12-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 13-AUG-1999; 990S-0148684.
PR 16-AUG-1999; 990S-0149175.
PR 17-AUG-1999; 990S-0149368.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149723.
PR 20-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0149930.
PR 23-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151303.
PR 31-AUG-1999; 990S-0151438.
PR 01-SEP-1999; 990S-0151930.
PR 07-SEP-1999; 990S-0152363.
PR 10-SEP-1999; 990S-0153070.
PR 13-SEP-1999; 990S-0153758.
PR 13-SEP-1999; 990S-0154018.
PR 16-SEP-1999; 990S-0154039.
PR 20-SEP-1999; 990S-0154779.
PR 22-SEP-1999; 990S-0155139.
PR 23-SEP-1999; 990S-0155486.
PR 24-SEP-1999; 990S-0155659.
PR 28-SEP-1999; 990S-0156458.
PR 29-SEP-1999; 990S-0156596.
PR 04-OCT-1999; 990S-0157117.
PR 05-OCT-1999; 990S-0157753.

PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161820.
 PR 28-OCT-1999; 99US-0161922.
 PR 28-OCT-1999; 99US-0161933.
 PR 29-OCT-1999; 99US-0162142.

Query Match 1.9%; Score 8; DB 21; Length 344;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 182 ESATEAER 189
 |||||
 DB 117 ESATEAER 124

RESULT 12
 ABG30935
 ID ABG30935 standard; Protein; 348 AA.
 XX
 AC ABG30935;

XX
 DT 21-OCT-2002 (first entry)
 XX

DE Human novel protein kinase 58848 protein.
 XX

XX Human: protein kinase 58848; protein kinase associated disorder;
 KW 58848-associated disorder; cellular disorder; proliferative disorder;
 KW differential disorder; cancer; leukaemia; hormonal disorder; diabetes;
 KW immune disorder; autoimmune disease; blood vessel disorder; hypertension;
 KW atherosclerosis; platelet disorder; cardiovascular disorder; ischaemia;
 KW cardiac hypertrophy; heart failure; neurological disorder; rickets;
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease; AIDS;
 KW acquired immunodeficiency syndrome; bone metabolism disorder; pain;
 KW osteoporosis; cirrhosis; haematopoietic neoplastic disorder; Hepatitis B;
 KW Hodgkin's disease; acute leukaemia; liver disorders; Gaucher's disease;
 KW viral disease; metabolic disorder; inflammation; hyperalgesia;
 KW chromosome mapping; tissue typing; forensic identification;
 KW pharmacogenomic profile.
 XX
 XX Homo sapiens.
 OS
 XX
 PN WO200255713-A2.
 XX
 PD 18-JUL-2002.
 XX
 PF 26-NOV-2001; 2001WO-US44346.

XX 08-DEC-2000; 2000US-254401P.
 XX
 XX (MILL-) MILLENNIUM PHARM INC.
 PA
 XX
 PI Kapeller-Libermann R, Acton S;
 XX
 DR WPI: 2002-590676/63.
 DR N-PSDB; ABR89200.

PT Novel human protein kinase polypeptide, designated 58848, useful for
 PT treating diseases including cellular, proliferative, bone metabolism,
 PT cardiovascular, neurological, and haematopoietic neoplastic disorders
 XX
 PS Claim 8; Page 99; 104pp; English.

CC The present invention relates to a new human protein kinase polypeptide
 CC designated 58848. The invention is useful for modulating 58848-mediated
 CC activities which are useful for developing diagnostic and therapeutic
 CC agents for protein kinase associated or other 58848-associated disorders
 CC such as cellular, proliferative and/or differential disorders e.g.
 CC cancer, leukaemia; hormonal disorders e.g. diabetes; immune disorders
 CC e.g. autoimmune disease; blood vessel disorders e.g. atherosclerosis;
 CC hypertension; platelet disorders; cardiovascular disorders e.g. cardiac
 CC hypertrophy; heart failure; neurological disorders e.g. ischaemia.
 CC Alzheimer's disease, Parkinson's disease, Huntington's disease, acquired
 CC immunodeficiency syndrome (AIDS); bone metabolism disorders e.g. rickets,
 CC osteoporosis, cirrhosis; haematopoietic neoplastic disorders e.g.
 CC Hodgkin's disease, acute leukaemia; liver disorders e.g. Gaucher's
 CC disease, viral diseases e.g. Hepatitis B; pain or metabolic disorder
 CC e.g. inflammation, hyperalgesia. The invention is also useful in
 CC chromosome mapping, tissue typing and in forensic identification of a
 CC biological sample. 58848 molecules are useful as markers of disorders or
 CC disease states, as markers of drug activity, or as markers of the
 CC pharmacogenomic profile of the subject. The present amino acid sequence
 CC represents the human novel protein kinase 58848 protein of the invention.
 XX

SQ Sequence 348 AA;

Query Match 1.9%; Score 8; DB 23; Length 348;
 Best Local Similarity 100.0%; Pred. No. 62;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 246 DELXEVR 253
 |||||
 DB 59 DELXEVR 66

RESULT 13
 AAE21721
 ID AAE21721 standard; Protein; 348 AA.
 XX
 AC AAE21721;

XX
 DT 16-JUL-2002 (first entry)
 XX

DE Human PKIN-16 protein.
 XX

XX Human: kinase; enzyme; PKIN-16 protein; immune system disorder; anaemia;
 KW acquired immune deficiency syndrome; thymic hypoplasia; Crohn's disease;
 KW asthma; neurological disorder; epilepsy; Charcot-Marie-Tooth disease;
 KW AIDS; seizures; cell proliferative disorder; cancer; adenocarcinoma;
 KW leukemia; lymphoma; melanoma; myeloma; sarcoma; developmental disorder;
 KW Down's syndrome; gene therapy; protein therapy; cytostatic.
 XX
 XX Homo sapiens.
 OS
 XX
 FH Key
 FT Domain
 FT Domain
 FT Domain
 FT Domain

Location/Qualifiers
 3..263
 /note="Protein kinase domain"
 62..315
 /note="Eukaryotic protein kinase domain"
 63..267

```
FT FT /note= "Protein kinase domain"
FT Domain 65..263
FT /note= "Protein kinase domain"
FT Domain 68..316
FT /note= "Protein kinase domain"
FT Domain 137..150
FT /note= "Tyrosine kinase catalytic domain"
FT Domain 173..191
FT /note= "Tyrosine kinase catalytic domain"
FT Domain 244..266
FT /note= "Tyrosine kinase catalytic domain"
PN WO200218557-A2.
XX
XX
XX 07-MAR-2002.
XX
XX 31-AUG-2001; 2001WO-US27219.
XX
XX 31-AUG-2000; 2000US-229873P.
XX 08-SEP-2000; 2000US-231357P.
XX 14-SEP-2000; 2000US-232654P.
XX 22-SEP-2000; 2000US-234902P.
XX 29-SEP-2000; 2000US-236499P.
XX 06-OCT-2000; 2000US-238389P.
XX 13-OCT-2000; 2000US-240542P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Bandman O, Nguyen DB, Walla NK, Hafalia AJA, Yao MG, Gandhi AR,
XX Gururajan R, Ding L, Patterson C, Yue H, Baughn MR, Tribouley CM,
XX Thornton M, Elliott VS, Lu Y, Ison CH, Au-Young J, Tang YT;
XX Azimzai Y, Burdill JD, Marcus GA, Zingler KA, Lu DAM, Lai PG;
XX Ramkumar J, Warren BA, Kearney L, Policky JL, Thangavelu K;
XX Burford N;
XX
XX WPI: 2002-329769/36.
XX N-PSDB: AAD34313.
XX
XX New human kinases, useful for diagnosing, treating or preventing immune
XX system disorders (e.g. Crohn's disease); neurological disorders (e.g.
XX epilepsy) or cell proliferative disorders (e.g. cancers such as
XX leukemia or lymphoma)
XX
XX Claim 71; Page 179-180; 218pp; English.
XX
XX The present invention relates to human kinases (PKIN) and polynucleotides
XX encoding such proteins. PKIN sequences of the invention are useful for
XX diagnosing, treating or preventing disorders associated with aberrant
XX expression of PKIN, particularly immune system disorders (e.g. acquired
XX immune deficiency syndrome (AIDS), thymic hypoplasia, Crohn's disease,
XX anaemia, asthma), neurological disorders (e.g. epilepsy, Charcot-Marie-
XX Tooth disease or seizures), cell proliferative disorders (e.g. cancers
XX such as adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma),
XX and developmental disorders (e.g. Down's syndrome). They are also used
XX in gene therapy and protein therapy. The present sequence is human
XX PKIN-16 protein.
XX
XX Sequence 348 AA:
XX
XX Query Match 1.9%; Score 8; DB 23; Length 348;
XX Best Local Similarity 100.0%; Pred. No. 62;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 246 DELVEEVR 253
XX |
XX Db 59 DELVEEVR 66
XX
XX RESULT 14
XX ID AAG90262 standard; Protein: 425 AA.
XX AC AAG90262;
```

```
XX XX
XX 26-SEP-2001 (first entry)
XX
XX C glutamicum protein fragment SEQ ID NO: 4016.
XX
XX DE Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis.
XX
XX KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis.
XX
XX OS Corynebacterium glutamicum.
XX
XX EP1108790-A2.
XX
XX
XX 20-JUN-2001.
XX
XX 18-DEC-2000; 2000EP-0127688.
XX
XX 16-DEC-1999; 99JP-0377484.
XX 07-APR-2000; 2000JP-0159162.
XX 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
XX Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI: 2001-376931/40.
XX N-PSDB: AAH65481.
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
XX mutation point of a gene, measuring expression of a gene, analysing
XX expression profile or pattern of a gene and identifying homologous gene
XX
XX Claim 17; SEQ ID NO: 4016; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
XX sequences from the Coryneform bacterium Corynebacterium glutamicum. These
XX are useful for identifying the mutation point of a gene derived from a
XX mutant of coryneform bacterium, measuring expression amount and
XX analysing the expression profile or expression pattern of a gene derived
XX from Coryneform bacterium, and identifying a homologue of a gene derived
XX from Coryneform bacterium. Coryneform bacteria are useful for producing
XX amino acids, nucleic acids, vitamins, saccharides and organic acids,
XX particularly L-lysine. The present sequence is a protein described
XX in the exemplification of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from the
XX European Patent Office.
XX
XX Sequence 425 AA:
XX
XX Query Match 1.9%; Score 8; DB 22; Length 425;
XX Best Local Similarity 100.0%; Pred. No. 74;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 322 TPSAPAAAS 329
XX |
XX Db 393 TPSAPAAAS 400
XX
XX RESULT 15
XX ID AAY95060 standard; Protein: 440 AA.
XX AC AAY95060;
XX
XX 23-JUN-2000 (first entry)
XX
XX Candida albicans polypeptide sequence # 28.
XX
XX DE Candida albicans infection; growth; survival; medicament; AIDS;
XX vulvovaginitis; immunocompromised patient; treat.
XX
XX KW
```

OS Candida albicans.

XX EP982401-A2.

PN 01-MAR-2000.

PD 23-DEC-1998; 98EP-0310694.

XX 14-AUG-1998; 98GB-0017796.

XX (JANC) JANSSEN PHARM NV.

XX Contreras RH, Nelissen B, De Backer MD, Luyten WHML, Viaene JE,

PI Logghe MG;

XX WPI; 2000-258614/23.

DR

XX Essential polypeptides isolated from Candida albicans, useful in the

PT treatment of diseases caused by C.albicans, especially in

PT immunocompromised subjects, e.g., AIDS patients -

XX

PS Claim 3; Page 75-76; 133pp; English.

XX

CC This sequence represents a polypeptide that is critical for the survival

CC and growth of Candida albicans. The C. albicans nucleic acid molecules

CC encoding the polypeptides of the invention may be used as probes and

CC primers for detecting homologous nucleic acid molecule sequences. The

CC polypeptides and nucleic acid molecules and compounds identified as

CC selectively modulating the expression of the polypeptides, may be used as

CC medicaments or for the preparation of a medicament to treat C.albicans

CC associated diseases, especially in AIDS patients and to treat

CC vulvovaginitis in otherwise healthy females. The use of the polypeptides

CC and polynucleotide sequences to treat C.albicans associated diseases has

CC fewer side effects and less toxicity than previously used methods such as

CC the use of amphotericin. This method is therefore especially suitable for

CC immunocompromised patients, such as AIDS patients.

XX

XX

Sequence 440 AA;

Query Match 1.9%; Score 8; DB 21; Length 440;

Best Local Similarity 100.0%; Pred. No. 76;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 365 ALYDYTAQ 372

Db 389 ALYDYTAQ 396

Search completed: July 28, 2003, 09:27:29

Job time : 88 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2003, 09:25:02 ; Search time 41 seconds
(without alignments)
973.415 Million cell updates/sec

Title: US-09-068-377C-1
Perfect score: 415
Sequence: 1 MMAQLQRFDAFWCRDFTAH.....WTVERNRGQRFVPSYLEKL 415

Scoring table: OLIGO
Gapop 60.0, Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size: 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: PIR_76:*

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	10	2.4	4639	1 A54794	dynein heavy chain
2	8	1.9	106	2 T06685	hypothetical prote
3	8	1.9	324	2 AB2772	inosine-uridine pr
4	8	1.9	378	2 A97552	chain a, crystal s
5	8	1.9	481	2 S62423	ATP-dependent RNA
6	8	1.9	536	2 T38210	scd2 protein - fls
7	8	1.9	541	2 T34701	hypothetical prote
8	8	1.9	586	2 C83262	hypothetical prote
9	8	1.9	633	2 T40124	hypothetical prote
10	8	1.9	633	2 B70954	kinesin-like motor
11	8	1.7	61	2 AE0689	ynbE protein precu
12	7	1.7	63	2 AE0664	probable lipoprote
13	7	1.7	105	2 H72736	hypothetical prote
14	7	1.7	133	2 AG3286	heat shock protein
15	7	1.7	135	2 D90348	hypothetical prote
16	7	1.7	148	1 A64158	glpe protein homol
17	7	1.7	148	1 D82623	conserved hypotet
18	7	1.7	148	2 A99348	hypothetical prote
19	7	1.7	168	2 G83428	hypothetical prote
20	7	1.7	174	2 E75532	conserved hypotet
21	7	1.7	176	2 T02217	NBS-LRR type resis
22	7	1.7	183	2 C97832	alpha-(1,3)-fucosy
23	7	1.7	185	2 G95110	translation initia
24	7	1.7	192	2 S66565	biotin carboxyl ca
25	7	1.7	192	2 JC4144	translation elonga
26	7	1.7	193	2 D72643	hypothetical prote
27	7	1.7	195	2 E97979	translation initia
28	7	1.7	203	2 S40235	rab24 protein - mo
29	7	1.7	214	2 D82839	hemolysin III prot

30	7	1.7	227	2 E64830	cytidylate kinase
31	7	1.7	227	2 A99753	cytidylate kinase
32	7	1.7	227	2 G85616	cytidylate kinase
33	7	1.7	227	2 AB0614	cytidylate kinase
34	7	1.7	229	2 C82141	cytidylate kinase
35	7	1.7	230	2 A10169	cytidylate kinase
36	7	1.7	246	2 T17668	hypothetical prote
37	7	1.7	251	2 F90331	partial transposas
38	7	1.7	263	2 G95198	hypothetical prote
39	7	1.7	263	2 C98065	hypothetical prote
40	7	1.7	270	2 D83072	conserved hypotet
41	7	1.7	285	2 D90429	first ORF - Sulfol
42	7	1.7	288	2 D90345	partial transpos
43	7	1.7	298	2 A70238	hypothetical prote
44	7	1.7	299	2 F95294	probable LysR-type
45	7	1.7	300	2 H84194	hypothetical prote

ALIGNMENTS

RESULT 1
A54794
dynein heavy chain, cytosolic - fruit fly (Drosophila melanogaster)
N:Contains: dynein ATPase (EC 3.6.4.2)
C:Species: Drosophila melanogaster
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Apr-2002
C:Accession: A54794
R:Li, M.; McGrail, M.; Serr, M.; Hays, T.S.
J. Cell Biol. 126, 1475-1494, 1994
A:Title: Drosophila cytoplasmic dynein, a microtubule motor that is asymmetrically lo
A:Reference number: A54794, MIMD:94375524, PMID:8089180
A:Accession: A54794
A:Molecule type: mRNA
A:Residues: 1-4639 <LIA>
A:Cross-References: GB:123195, NID:g349668, PIDN:AAA60323.1, PID:g349669
C:Genetics:
A:Gene: FlyBase:Dhnc4C
A:Cross-References: FlyBase:Fbgn0010349
C:Superfamily: dynein heavy chain, cytosolic
C:Keywords: ATP; heterotetramer; hydrolase; microtubule binding; nucleotide binding;
F:1895-1902/Region: nucleotide-binding motif A (P-loop)
F:2210-2217/Region: nucleotide-binding motif A (P-loop)
F:2580-2587/Region: nucleotide-binding motif A (P-loop)
F:2922-2929/Region: nucleotide-binding motif A (P-loop)
F:1901/Binding site: ATP (Lys) #status predicted
F:2216/Binding site: ATP (Lys) #status predicted
F:2586/Binding site: ATP (Lys) #status predicted
F:2928/Binding site: ATP (Lys) #status predicted

Query Match 2.4%; Score 10; DB 1; Length 4639;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 98 LREELSLLE 107
DB 3398 LREELSLLE 3407

RESULT 2
T06685
hypothetical protein T17F15.150 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 22-Oct-1999
C:Accession: T06685
R:Quetier, F.; Choisne, N.; Robert, C.; Brothier, P.; Wincker, P.; Cattoioco, L.; Art
submitted to the Protein Sequence Database, April 1999
A:Reference number: Z15793
A:Accession: T06685
A:Molecule type: DNA
A:Residues: 1-106 <QUE>
A:Cross-References: EMBL:AL049658; GSPDB:GN00061; ATSP:T17F15.150
A:Experimental source: cultivar Columbia; BAC clone T17F15

C:Genetics:
A:Gene: ATSP:T17F15.150
A:Map position: 3
A:Introns: 40/3

Query Match 1.9%; Score 8; DB 2; Length 106;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 321 TTPSAPPA 328
Db 32 TTPSAPPA 39

RESULT 3
AB2772
Inosine-uridine preferring nucleoside hydrolase [imported] - Agrobacterium tumefaciens
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AB2772
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Moo, I.; Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kuttyavin, T.; Levy, R.; Li, M.; McClell
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Kresspan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:21608550; PMID:11743193
A:Accession: AB2772
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1324 <KUR>
A:Cross-references: GB:AE008668; PIDN:AAL42592.1; PID:g17740018; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: iunH
A:Map position: circular chromosome
C:Superfamily: yaaf protein

Query Match 1.9%; Score 8; DB 2; Length 324;
Best Local Similarity 100.0%; Pred. No. 9.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 94 LALALRE 101
Db 137 LALALRE 144

RESULT 4
A97552
Chain a, crystal structure of nucleoside hydrolase from leishmania major [imported] - Ag
C:Species: Agrobacterium tumefaciens
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C:Accession: A97552
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollem, C.; Allinger, M.; Doughy, D.; Scott, C.; Lapps, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:21608551; PMID:11743194
A:Accession: A97552
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-378 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK67370.1; PID:g15156676; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_2923
A:Map position: circular chromosome
C:Superfamily: yaaf protein

Query Match 1.9%; Score 8; DB 2; Length 378;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 94 LALALRE 101
Db 191 LALALRE 198

RESULT 5
S62423
ATP-dependent RNA helicase - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 16-May-1996 #sequence_revision 13-Mar-1997 #text_change 02-Feb-2001
C:Accession: T38183; S62423
R:Iye, G.; Church, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, October 1995
A:Reference number: Z21776
A:Accession: T38183
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-481 <LY2>
A:Cross-references: EMBL:Z54285; PIDN:CAA91073.1; PID:g1008436; GSPDB:GN00066; SPDB:S
A:Experimental source: strain 972h-; cosmid c22F3
C:Genetics:
A:Gene: SPAC22F3.08c
A:Map position: 1L
C:Keywords: ATP; nucleotide binding; P-loop
F:90-97/Region: nucleotide-binding motif A (P-loop)
F:192-197/Region: nucleotide-binding motif B
F:196-199/Region: DEAD motif

Query Match 1.9%; Score 8; DB 2; Length 481;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 RQEQRRK 118
Db 450 RQEQRRK 457

RESULT 6
T38210
scd2 protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Aug-2002
C:Accession: T38210; T52491
R:Grishchuk, K.; McIntosh, J.R.; Devlin, K.; Church, C.; Barrell, B.G.; Rajandream,
submitted to the EMBL Data Library, February 1996
A:Reference number: Z21778
A:Accession: T38210
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-536 <DEV>
A:Cross-references: EMBL:Z69730; PIDN:CAA93608.1; GSPDB:GN00066; SPDB:SPAC22H10.07
A:Experimental source: strain 972h(-); cosmid c22H10
R:Chang, E.C.; Barr, M.; Wang, Y.; Jung, V.; Xu, H.P.; Wiegler, M.H.
Cell 79, 131-141, 1994
A:Title: Cooperative interaction of S. pombe proteins required for mating and morphog
A:Reference number: Z09915
A:Accession: T52491
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-536 <CHA>
A:Cross-references: EMBL:U12539; PIDN:AAA50557.1
C:Genetics:
A:Gene: SPAC22H10.07
A:Map position: 1
A:Introns: 3/3
A:Note: scd2
C:Function:
A:Description: required for normal morphology and mating
C:Superfamily: fission yeast scd2 protein; SH3 homology

Query Match 1.9%; Score 8; DB 2; Length 536;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 364 RALDYTA 371
|||||||
DB 30 RALDYTA 37

RESULT 7

hypothetical protein SCIC3.09 SCIC3.09 - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C:Accession: T34701
R:Oliver, K.; Harris, D.; Parkhill, J.; Barrrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, May 1998
A:Reference number: Z21554
A:Accession: T34701
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-541 <OLI>
A:Cross-references: EMBL:AL023702; PIDN:CAA19233.1; GSPDB:GN00070; SCOEDB:SCIC3.09
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCIC3.09

Query Match 1.9%; Score 8; DB 2; Length 541;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 LQRLDGR 33
|||||||
DB 494 LQRLDGR 501

RESULT 8

hypothetical protein PA3074 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: C63262
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Badian, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapid, K.; Lam,
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: C63262
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-586 <STO>
A:Cross-references: GB:AE004731; GB:AE004091; NID:g9949171; PIDN:AAG06462.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA3074

Query Match 1.9%; Score 8; DB 2; Length 586;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 39 VEELLROR 46
|||||||
DB 452 VEELLROR 459

RESULT 9

kinesin-like motor protein SPBC2D10.21c - fission yeast (Schizosaccharomyces pombe)
T40124
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
C:Accession: T40124; T39477
R:Wood, V.; Rajandream, M.A.; Barrrell, B.G.; Taylor, K.; Harris, D.
submitted to the EMBL Data Library, September 1998
A:Reference number: Z21906
A:Accession: T40124

A:Molecule type: DNA
A:Residues: 173-633 <MOO>
A:Cross-references: EMBL:AL031788; PIDN:CAA21179.1; GSPDB:GN00067; SPDB:SPBC2D10.21c
A:Experimental source: strain 972h; cosmid c2D10
R:Lyne, M.; Rajandream, M.A.; Barrrell, B.G.; Lucas, M.; Gallardin, C.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z21858
A:Accession: T39477
A:Molecule type: DNA

A:Residues: 1-297 <LYN>
A:Cross-references: EMBL:AL031349; PIDN:CAA20476.1; GSPDB:GN00067; SPDB:SPBC15D4.01c
A:Experimental source: strain 972h; cosmid c15D4
C:Genetics:
A:Gene: SPDB:SPBC2D10.21c; SPDB:SPBC15D4.01c
A:Map position: 2

Query Match 1.9%; Score 8; DB 2; Length 633;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 100 EELRSLDE 107
|||||||
DB 526 EELRSLDE 533

RESULT 10

hypothetical protein RV1251c - Mycobacterium tuberculosis (strain H37RV)
B70954
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: B70954
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:9825987; PMID:9634230
A:Accession: B70954
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1139 <COL>
A:Cross-references: GB:AL021006; GB:AL123456; NID:g3242277; PIDN:CAA15907.1; PID:e121
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV1251c

Query Match 1.9%; Score 8; DB 2; Length 1139;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 25 LQRLDGR 32
|||||||
DB 191 LQRLDGR 198

RESULT 11

ynde protein precursor - Escherichia coli (strain K-12)
A64889
C:Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: A64889
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: A64889
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-61 <BLAT>
A:Cross-references: GB:AE000235; GB:U00096; NID:g1787643; PIDN:AAC74464.1; PID:g17876
A:Experimental source: strain K-12, substrain MG1655

C:Genetics:
A:Gene: ynbE
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-61/Product: ynbE protein #status predicted <MAT>

Query Match 1.7%; Score 7; DB 2; Length 61;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 KDVEELL 43
|||||
DB 48 KDVEELL 54

RESULT 12

AE0664
Probable lipoprotein STY1424 [imported] - *Salmonella enterica* subsp. *enterica* serovar Typhimurium
C:Species: *Salmonella enterica* subsp. *enterica* serovar Typhimurium
A:Note: this species has also been called *Salmonella typhimurium*
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AE0664

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moulie, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhimurium
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AE0664
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-63 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD01686.1; PID:g16502537; GSPDB:GN00176
C:Genetics:
A:Gene: STY1424

Query Match 1.7%; Score 7; DB 2; Length 63;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 KDVEELL 43
|||||
DB 50 KDVEELL 56

RESULT 13

H72736
Hypothetical protein APE0430 - *Aeropyrum pernix* (strain K1)
C:Species: *Aeropyrum pernix*
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: H72736

R:Kawarabayashi, Y.; Hirao, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, *Aeropyrum pernix*
A:Reference number: A72450; MUID:99310339; PMID:10382966
A:Accession: H72736
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-105 <KAW>
A:Cross-references: DBJ:AP000059; NID:g5103911; PIDN:BAA79388.1; PID:g5104072
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE0430
C:Superfamily: *Aeropyrum pernix* hypothetical protein APE0430

Query Match 1.7%; Score 7; DB 2; Length 105;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 VLLORLL 30
|||||
DB 32 VLLORLL 38

RESULT 14

AG3286

Heat shock protein 15 [imported] - *Brucella melitensis* (strain 16M)

C:Species: *Brucella melitensis*

C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002

C:Accession: AG3286

R:DelVecchio, V.G.; Kapetral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanov, M.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A:Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*

A:Reference number: AD3252; PMID:11756688

A:Accession: AG3286

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-133 <KUR>

A:Cross-references: GB:AE008917; PIDN:AL51458.1; PID:g17982168; GSPDB:GN00190

A:Experimental source: strain 16M

C:Genetics:
A:Gene: BME10277

A:Map position: 1

Query Match 1.7%; Score 7; DB 2; Length 133;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 326 PAASTET 332
|||||
DB 102 PAASTET 108

RESULT 15

D90348

Hypothetical protein SSO1853 [imported] - *Sulfolobus solfataricus* transposon ISC1359

C:Species: *Sulfolobus solfataricus*

C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001

C:Accession: D90348

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aways, M.J.; Ch Jung, I.; Ueffing, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder

submitted to Genbank, April 2001

A:Description: *Sulfolobus solfataricus* complete genome.

A:Reference number: A99139

A:Accession: D90348

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-135 <KUR>

A:Cross-references: GB:AE006641; NID:g13815107; PIDN:AAK42043.1; GSPDB:GN00155

C:Genetics:
A:Gene: SSO1853

Query Match 1.7%; Score 7; DB 2; Length 135;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 103 RSLTEFR 109
|||||
DB 62 RSLTEFR 68

Search completed: July 28, 2003, 09:30:36
Job time: 44 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2003, 09:18:22 ; Search time 23 Seconds

(without alignments)
848.525 Million cell updates/sec

Title: US-09-068-377c-1

Perfect score: 415

Sequence: 1 MMAQLQFRDAFCRCDFTAHT.....MTVERNGRGVEVGSYLEKL 415

Scoring table: OLIGO

Searched: 127863 seqs, 47026705 residues

Word size: 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	2.4	4639	1	DYHC_DROME
2	8	1.9	481	1	YA88_SCHPO
3	8	1.9	536	1	SCD2_SCHPO
4	8	1.9	1134	1	ANSI_HUMAN
5	8	1.9	1150	1	ANSI_MOUSE
6	7	1.7	61	1	YNBE_ECOLI
7	7	1.7	144	1	P532_MOUSE
8	7	1.7	148	1	Y1BN_HAEIN
9	7	1.7	176	1	IF3_STRAB
10	7	1.7	176	1	IF3_STRAB
11	7	1.7	182	1	RR4_HYMLI
12	7	1.7	185	1	IF3_STRAB
13	7	1.7	203	1	RB24_HUMAN
14	7	1.7	203	1	RB24_MOUSE
15	7	1.7	225	1	KCY_VIBCH
16	7	1.7	227	1	KCY_ECOLI
17	7	1.7	227	1	KCY_SALTY
18	7	1.7	227	1	KCY_SALTY
19	7	1.7	227	1	KCY_SALTY
20	7	1.7	230	1	KCY_YERPE
21	7	1.7	299	1	NUSG_STRVG
22	7	1.7	306	1	P2A1_ARATH
23	7	1.7	306	1	P2A2_ARATH
24	7	1.7	307	1	P2A5_ARATH
25	7	1.7	309	1	NUSG_STRGB
26	7	1.7	313	1	RLAO_NEUCR
27	7	1.7	318	1	CPWM_AERPE
28	7	1.7	326	1	YGGF_ECOLI
29	7	1.7	350	1	AROG_ECOLI
30	7	1.7	352	1	ALGP_PSEAE
31	7	1.7	353	1	RFC2_YEAST
32	7	1.7	357	1	HRCU_RALSO
33	7	1.7	360	1	AROC_ECOL6

ALIGNMENTS

RESULT 1	DYHC_DROME	STANDARD	PRT: 4639 AA.
AC	P37276		
DT	01-OCT-1994 (Rel. 30, Created)		
DT	01-OCT-1994 (Rel. 30, Last sequence update)		
DT	15-SEP-2003 (Rel. 42, Last annotation update)		
DE	Dynein heavy chain, cytosolic (DYHC).		
GN	CDHC OR DHC64C.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;		
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=94375524; PubMed=8089180;		
RA	Li M., McGrail M., Serr M., Hays T.S.;		
RT	"Drosophila cytoplasmic dynein, a microtubule motor that is		
RL	asymmetrically localized in the oocyte.";		
RL	J. Cell Biol. 126:1475-1494(1994).		
RP	SEQUENCE OF 1877-1998 FROM N.A.		
RX	MEDLINE=94243034; PubMed=8186464;		
RA	Rasmussen K., Serr M., Geppert J., Gibbons I., Hays T.S.;		
RT	"A family of dynein genes in Drosophila melanogaster.";		
RL	Mol. Biol. Cell 5:45-55(1994).		
CC	-1- FUNCTION: Cytoplasmic dynein acts as a motor for the intracellular		
CC	retrograde motility of vesicles and organelles along microtubules.		
CC	Dynein has ATPase activity; the force-producing power stroke is		
CC	thought to occur on release of ADP.		
CC	-1- SUBUNIT: Consists of at least two heavy chains and a number of		
CC	intermediate and light chains.		
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic.		
CC	-1- DOMAIN: Dynein heavy chains probably consist of an N-terminal stem		
CC	(which binds cargo and interacts with other dynein components),		
CC	and the head or motor domain. The motor contains six tandemly-		
CC	linked AAA domains in the head, which form a ring. A stalk-like		
CC	structure (formed by two of the coiled coil domains) protrudes		
CC	between AAA 4 and AAA 5 and terminates in a microtubule-binding		
CC	site. A seventh domain may also contribute to this ring; it is not		
CC	clear whether the N-terminus or the C-terminus forms this extra		
CC	domain. There are four well-conserved and two non-conserved ATPase		
CC	sites, one per AAA domain. Probably only one of these (within AAA		
CC	1) actually hydrolyzes ATP, the others may serve a regulatory		
CC	function.		
CC	-1- SIMILARITY: Belongs to the dynein heavy chain family.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		

34	7	1.7	360	1	AROC_ECOLI	P12008 escherichia
35	7	1.7	360	1	AROC_YERPE	O8ad41 yersinia pe
36	7	1.7	362	1	B10B_SYNY3	P73538 synechocyst
37	7	1.7	379	1	MUN_DROME	O9v576 drosophila
38	7	1.7	408	1	FEK_THEVO	O97bc6 thermoplasm
39	7	1.7	413	1	DXR_MYCTU	O10798 mycobacteri
40	7	1.7	443	1	G6PI_STAAM	O99vvc staphylococ
41	7	1.7	443	1	G6PI_STAAM	O8nxfl staphylococ
42	7	1.7	460	1	PPAN_DROME	O9vdes drosophila
43	7	1.7	468	1	K1CP_MOUSE	O922K1 mus musculu
44	7	1.7	504	1	BLK_HUMAN	P51451 homo sapien
45	7	1.7	508	1	CP2B_HUMAN	O15528 homo sapien

```

DR EMBL: L23195; AAA60323.1;
DR EMBL: L25122; AAA28492.1;
DR PIR: A54794; A54794;
DR FlyBase: FBgn0010349; Dhc64C.
DR GO: GO:0045169; C:fusome; NAS.
DR GO: GO:0008567; F:dynamin-ATPase activity; IDA.
DR GO: GO:0003777; F:microtubule motor activity; IDA.
DR GO: GO:0007098; P:centrosome cycle; IMP.
DR GO: GO:0045478; P:fusome organization and biogenesis; NAS.
DR GO: GO:0007018; P:microtubule-based movement; IDA.
DR GO: GO:0007294; P:ooocyte determination; IMP.
DR GO: GO:0007292; P:oogenesis; IMP.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR004273; Dynein_heavy.
DR Pfam: PF03028; Dynein_heavy; 1.
DR SMART: SM00382; AAA; 4.
KW Motor protein; Dynein; Microtubules; ATP-binding; Repeat; Coiled coil.
FT DOMAIN 1 1856 STEM (BY SIMILARITY).
FT DOMAIN 1857 2084 AAA 1 (BY SIMILARITY).
FT DOMAIN 2166 2437 AAA 2 (BY SIMILARITY).
FT DOMAIN 2541 2790 AAA 3 (BY SIMILARITY).
FT DOMAIN 2884 3153 AAA 4 (BY SIMILARITY).
FT DOMAIN 3189 3478 STALK (BY SIMILARITY).
FT DOMAIN 3539 3768 AAA 5 (BY SIMILARITY).
FT DOMAIN 3989 4205 AAA 6 (BY SIMILARITY).
FT DOMAIN 530 565 COILED COIL (POTENTIAL).
FT DOMAIN 774 794 COILED COIL (POTENTIAL).
FT DOMAIN 1264 1368 COILED COIL (POTENTIAL).
FT DOMAIN 3189 3261 COILED COIL (POTENTIAL).
FT DOMAIN 3382 3478 COILED COIL (POTENTIAL).
FT DOMAIN 3723 3782 COILED COIL (POTENTIAL).
FT NP_BIND 1895 1902 ATP (POTENTIAL).
FT NP_BIND 2210 2217 ATP (POTENTIAL).
FT NP_BIND 2580 2587 ATP (POTENTIAL).
FT NP_BIND 2922 2928 ATP (POTENTIAL).
SQ SEQUENCE 4639 AA; 530152 MW; 057A7D8800CCD07E CRC64;

Query Match 2.4%; Score 10; DB 1; Length 4639;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 LRELRSLER 107
DB 3398 LRELRSLER 3407

RESULT 2
YAB8_SCHPO STANDARD; PRT; 481 AA.
ID YAB8_SCHPO
AC 009775;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative ATP-dependent RNA helicase C22F3.08c.
GN SPAC22F3.08c.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OC NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown R., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Mount S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver P., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,

```

```

RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tiley A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Roben J., Grymopiez B.,
RA Wajsbjerg S., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehnrich H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Mamult R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
RL The genome sequence of Schizosaccharomyces pombe.
RL Nature 415:871-880(2002).
CC -1- FUNCTION: PUTATIVE ATP-DEPENDENT RNA HELICASE.
CC -1- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z54285; CA91073.1;
DR PIR: T38183; S62423.
DR HSSP: Q58083; 1HV8.
DR GeneDB: spombe: SPAC22F3.08c;
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR000629; DEAD_box.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF00270; DEAD; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELICC; 1.
DR PROSITE: PS00039; DEAD_ATP_HELICASE; 1.
KW Hypothetical protein; ATP-binding; RNA-binding; Helicase.
FT NP_BIND 90 97 ATP (POTENTIAL).
FT SITE 196 199 DEAD BOX.
SQ SEQUENCE 481 AA; 54026 MW; 872ECAPC106E110A CRC64;

Query Match 1.9%; Score 8; DB 1; Length 481;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 RQKQRKK 118
DB 450 RQKQRKK 457

RESULT 3
SCD2_SCHPO STANDARD; PRT; 536 AA.
ID SCD2_SCHPO
AC P40956;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein scd2/ral3.
GN SCD2 OR RAL3 OR SPAC22H10.07.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OC NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SP870;
RX MEDLINE=95007774; PubMed=7923372;
RA Chang E.C., Barr M., Wang Y., Jung V., Xu H.-P., Wiggler M.H.;

```

*Cooperative interaction of S. pombe proteins required for mating and morphogenesis.";
 RT Cell 79:131-141(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972;
 RX MEDLINE-21648401; PubMed-11859360;
 RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holyoak S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones I., Jones M., Leather S., McDonald S., O'Connell J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., O'Neill C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Stevens K., Sharp S.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
 RA Welter J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Usero D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe";
 RL Nature 415:871-880(2002).
 CC -I- FUNCTION: REQUIRED FOR MATING AND MORPHOGENESIS. INTERACTS
 DIRECTLY WITH SCG1 AND WITH CDC42. MAY BRIDGE AND FACILITATE SCG1
 AND CDC42 INTERACTIONS.
 CC -I- SUBUNIT: SCG1, SCG2, CDC42, AND RAS1, IN ITS GTP-BOUND STATE, ACT
 COOPERATIVELY TO FORM A PROTEIN COMPLEX.
 CC -I- SIMILARITY: Contains 2 SH3 domains.
 CC -I- SIMILARITY: SPRONG, TO YEAST BEML
 CC -I- SIMILARITY: Contains 1 phox homology (PX) domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: 012539; AAA50557.1; -
 CC EMBL: 269730; CAA93608.1; -
 CC PIR: T38210; T38210.
 DR HSP: 060631; IGBO.
 DR GenDB: SPombe; SPAC22H10.07; -
 DR InterPro: IPR000270; OPR_PBI.
 DR InterPro: IPR001683; PX.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00564; PBI; 1.
 DR Pfam: PF00787; PX; 1.
 DR Pfam: PF00018; SH3; 2.
 DR PRINTS: PR00452; SH3DOMAIN.
 DR ProDom: PD000066; SH3; 1.
 DR SMART: SM00666; PBI; 1.
 DR SMART: SM00312; PX; 1.
 DR SMART: SM00326; SH3; 2.
 DR PROSITE: PS50195; PX; 1.
 DR PROSITE: PS50002; SH3; 2.
 KW SH3 domain; Repeat.
 FT DOMAIN 24 86 SH3 1.
 FT DOMAIN 123 185 SH3 2.
 FT DOMAIN 293 413 PX.
 SQ SEQUENCE 536 AA; 60042 MW; 7B86BA8CB32D1762 CRC64;

Query Match 1.9%; Score 8; DB 1; Length 536;
 Best Local Similarity 100.0%; Pred. No. 8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 364 RALDYXTA 371
 Db 30 RALDYXTA 37
 |||||
 RESULT 4
 ID ANSL_HUMAN STANDARD; PRT; 1134 AA.
 AC 092625;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DE Ankyrin repeat and SAM domain containing protein 1.
 GN ANKSL OR KIAA0229.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Bone marrow;
 RX MEDLINE-97191544; PubMed-9039502;
 RA Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayashi Y.,
 RA Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;
 RT "Prediction of the coding sequences of unidentified human genes. VI.
 RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
 RT analysis of cDNA clones from cell line KG-1 and brain.";
 RL DNA Res. 3:321-329(1996).
 CC -I- SIMILARITY: Contains 6 ANK repeats.
 CC -I- SIMILARITY: Contains 1 PID domain.
 CC -I- SIMILARITY: Contains 2 SAM domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D86982; BAA13218.1; ALT_INT.
 DR HSP: 002410; IAQC.
 DR Genew: HGNC:20961; ANKSL.
 DR InterPro: IPR002110; ANKS.
 DR InterPro: IPR006020; PTB_PID.
 DR InterPro: IPR001660; SAM.
 DR Pfam: PF00023; ank; 7.
 DR Pfam: PF00640; PID; 1.
 DR Pfam: PF00536; SAM; 2.
 DR PRINTS: PR01415; ANKYRIN.
 DR SMART: SM00248; ANK; 6.
 DR SMART: SM00462; PTB; 1.
 DR SMART: SM00454; SAM; 2.
 DR PROSITE: PS50086; ANK_REPEAT; 6.
 DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE: PS01179; PID; 1.
 DR PROSITE: PS50105; SAM_DOMAIN; 2.
 KW ANK repeat; Repeat.
 FT DOMAIN 1 55 GLY-RICH.
 FT REPEAT 79 108 ANK 1.
 FT REPEAT 112 141 ANK 2.
 FT REPEAT 148 177 ANK 3.
 FT REPEAT 181 210 ANK 4.
 FT REPEAT 214 243 ANK 5.
 FT REPEAT 246 275 ANK 6.
 FT REPEAT 296 327 SAM 1.
 FT DOMAIN 696 762 SAM 1.
 FT DOMAIN 770 837 SAM 2.
 FT DOMAIN 936 1091 PID.

SQL	SEQUENCE	1134 AA	123081 MW	21B7171BC6A800BA CRC64
Query Match		1.9%	Score 8:	DB 1; Length 1134;
Best Local Similarity		100.0%	Pred. No. 15;	
Matches	8; Conservative	0;	Mismatches	0; Indels
		0;	Gaps	0;
Oy	311 FSGLLHGS 318			
Db	499 FSGLLHGS 506			
RESULT 5				
ANSI_MOUSE				
ID	ANSI_MOUSE	STANDARD:	PRT:	1150 AA.
AC	P59672;			
DT	15-SEP-2003 (Rel. 42, Created)			
DT	15-SEP-2003 (Rel. 42, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Ankyrin repeat and SAM domain containing protein 1.			
GN	ANK1.			
OS	Mus musculus (Mouse).			
OC	Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Limb;			
RC	MEDLINE=22388257; PubMed=12477932;			
RA	Strausberg R.D., Feingold E.A., Grouse L.H., Derge J.G.,			
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,			
RA	Altshul S.F., Zeeberg B., Buelow K.H., Scheet C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Ditschenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,			
RA	Brownstein M.J., Ussid T.B., Toshiyuki S., Carlinici P., Prange C.,			
RA	Rah S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,			
RA	Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA	Richards S., Worley R.C., Hale S., Garcia A.M., Gay L.J., Holyk S.W.,			
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,			
RA	Faney J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,			
RA	Whiting R.W., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,			
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,			
RA	Butterfield V.S.N., Krzywinski M.T., Skalski U., Smalusz D.E.,			
RA	Schmeh A., Schin J.E., Jones S.J.M., Maira M.A.,			
RT	"Generation and initial analysis of more than 15,000 full-length			
RT	human and mouse cDNA sequences."			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).			
CC	-1- SIMILARITY: Contains 6 ANK repeats.			
CC	-1- SIMILARITY: Contains 1 PID domain.			
CC	-1- SIMILARITY: Contains 2 SAM domains.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; BC050847; AAH50847.1; -			
DR	SMART; SM00248; ANK. 6.			
DR	SMART; SM00462; PTB. 1.			
DR	SMART; SM00454; SAM. 2.			
DR	PROSITE; PS50297; ANK_REPEAT; 1.			
DR	PROSITE; PS50088; ANK_REPEAT; 5.			
DR	PROSITE; PS01179; PID. 1.			
DR	PROSITE; PS0105; SAM_DOMAIN; 2.			
DM	ANK repeat; Repeat.			
FT	DOMAIN 31 51 GLY-RICH.			
FT	REPEAT 75 104 ANK 1.			
FT	REPEAT 108 137 ANK 2.			
FT	REPEAT 165 194 ANK 3.			

Query Match	Best Local	Similarity	1.9%;	Score 8;	DB 1;	Length 1150;
Matches	8;	Conservative	100.0%;	Pred. 0.15;	Mismatches	0; Indels 0; Gaps 0;
FT REPEAT	198	227		ANK 4.		
FT REPEAT	231	260		ANK 5.		
FT REPEAT	263	292		ANK 6.		
FT DOMAIN	712	778		SAM 1.		
FT DOMAIN	786	853		SAM 2.		
FT DOMAIN	952	1107		PID		
SEQ SEQUENCE	1150 AA;	125241 MW;	551AF44AD1DB02D1B	CR64;		
QY	311	FSGLHGS 318				
DB	516	FSGLHGS 523				
RESULT 6						
YMBE_ECOLI	STANDARD:	PRT;	61 AA.			
AC P76075;						
DT 15-JUL-1998 (Rel. 36, Created)						
DT 15-JUL-1998 (Rel. 36, Last sequence update)						
DT 28-FEB-2003 (Rel. 41, Last annotation update)						
DE Hypothetical protein ymbE.						
GN YMBE OR B1382 OR C1829 OR SF1816.						
OS Escherichia coli,						
OS Escherichia coli 06, and						
OS Shigella flexneri.						
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;						
OC Enterobacteriaceae; Escherichia.						
OX NCBI_TaxID=562, 217992, 623;						
[1]						
SEQUENCE FROM N.A.						
RP SPECIES=E.coli; STRAIN=K12 / MG1655;						
RX MEDLINE=97426617; PubMed=9278503;						
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,						
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,						
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,						
RA Mau B., Shao Y.;						
RT "The complete genome sequence of Escherichia coli K-12.";						
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(1997).						
[2]						
SEQUENCE FROM N.A.						
RP SPECIES=E.coli; STRAIN=O6:H1 / CFT073 / ATCC 700928;						
RX MEDLINE=22274206; PubMed=12384590;						
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,						
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,						
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,						
RA Chang H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,						
RA Yu J.;						
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity						
RT through comparison with genomes of Escherichia coli K12 and O157.";						
RL Nucleic Acids Res. 30:4432-4441(2002).						
[3]						
SEQUENCE FROM N.A.						
RP SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;						
RX MEDLINE=22272406; PubMed=12384590;						
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,						
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,						
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,						
RA Chang H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,						
RA Yu J.;						
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity						
RT through comparison with genomes of Escherichia coli K12 and O157.";						
RL Nucleic Acids Res. 30:4432-4441(2002).						
CC This SWISS-PROT entry is copyright. It is produced through a collaboration						
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -						
CC the European Bioinformatics Institute. There are no restrictions on its						
CC use by non-profit institutions as long as its content is in no way						
CC modified and this statement is not removed. Usage by and for commercial						
CC entities requires a license agreement (See http://www.isb-sib.ch/annouce/						
CC or send an email to license@isb-sib.ch).						

```

CC -----
DR EMBL: AE000235; AAC74464.1; -
DR EMBL: AE016760; AAN80293.1; -
DR EMBL: AE015201; AANA3382.1; -
DR PIR: A64889; A64889.
DR Ecogene: EG14298; ynbE.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 61 AA; 6840 MW; CBE30197B8C5233E CRC64;

Query Match
Best Local Similarity 1.7%; Score 7; DB 1; Length 61;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 37 KDVEEL 43
DB 48 KDVEEL 54

RESULT 7
P532_MOUSE STANDARD; PRT; 144 AA.
ID P532_MOUSE
AC 062415:
DT 15-JUL-1998 (Rel. 36, Created).
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Tumor suppressor p53-binding protein 2 (p53-binding protein 2) (53BP2)
DE (SH3P1 protein) (Fragment).
GN TP53BP2 OR TRP53BP2 OR P53BP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:98294438; PubMed:9630982;
RA Sparks A.B., Hoffman N.G., McConnell S.J., Fowlkes D.M., Kay B.K.;
RT "Cloning of ligand targets: systematic isolation of SH3 domain-
   containing proteins.";
RT Nat. Biotechnol. 14:741-744(1996).
CC -1- SUBUNIT: IMPEDES CELL CYCLE PROGRESSION AT G2/M (BY SIMILARITY).
CC -1- SUBUNIT: BINDS TO THE CENTRAL DOMAIN OF P53 AS WELL AS TO BCL-2.
CC -1- INTERACTS WITH PROTEIN PHOSPHATASE 1 (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- DOMAIN: SPECIFIC INTERACTIONS WITH EITHER BCL-2 OR P53 REQUIRE THE
   ANKYRIN REPEATS AND THE SH3 DOMAIN (BY SIMILARITY).
CC -1- SIMILARITY: Contains at least 1 ANK repeat.
CC -1- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U58881; AAC52638.1; -
DR HSSP: 013625; lycS.
DR MGD: MGI:1336199; Trp53bp2.
DR InterPro: IPR002110; ANK.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00023; ank; 1.
DR Pfam: PF00018; SH3; 1.
DR PRINTS: PR00452; SH3DOMAIN.
DR PRODOM: PD000066; SH3; 1.
DR SMART: SM00248; ANK; 1.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS50088; ANK_REPEAT; 1.
DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE: PS50002; SH3; 1.
KW SH3 domain; Repeat; ANK repeat.
FT NON_TER 1
FT REPEAT 7 36 ANK.

```

```

FT DOMAIN 74 138 SH3.
SQ NON_TER 144 144
SQ SEQUENCE 144 AA; 16249 MW; 6D688C9C07B20FA3 CRC64;

Query Match
Best Local Similarity 1.7%; Score 7; DB 1; Length 144;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 371 AONSD 377
DB 86 AONSD 92

RESULT 8
YIBN_HAEIN STANDARD; PRT; 148 AA.
ID YIBN_HAEIN
AC P44854;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein HI0744.
GN HI0744.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxId=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE-95350630; PubMed-7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Ravelle A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-T., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
   Rd.";
RT Science 269:496-512(1995).
CC -1- SIMILARITY: Contains 1 rhodanese domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32758; AAC22402.1; -
DR PIR: A64158; A64158.
DR TIGR: HI0744; -
DR InterPro: IPR001763; Rhodanese-like.
DR Pfam: PF00581; Rhodanese; 1.
DR SMART: SM00450; RHODANSE; 1.
DR PROSITE: PS50206; RHODANSE_3; 1.
KW Hypothetical protein; Complete proteome.
FT DOMAIN 54 148 RHODANSE.
SQ SEQUENCE 148 AA; 16537 MW; 6CDC1E29CADA06B CRC64;

Query Match
Best Local Similarity 1.7%; Score 7; DB 1; Length 148;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 102 LRSLEP 108
DB 63 LRSLEP 69

RESULT 9

```

IF3-STRA3
ID IF3-STRA3 STANDARD: PRT: 176 AA.
AC O8E4E7: O8DYT9:
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE Translation initiation factor IF-3.
GN INFC OR GBS1454 OR SAG1384.
OS Streptococcus agalactiae (serotype III), and
OC Streptococcus agalactiae (serotype V).
OC Bacteria: Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216495, 216466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NEM316 / Serotype III;
RX MEDLINE=22242508; PubMed=12354221;
RA Glaser P., Rusniok C., Buchrieser C., Chevallier F., Frangoul L.,
RA Madef T., Zouine M., Couve E., Lallou L., Poyart C., Tlieu-Cuot P.,
RA Kunst F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease."
RL Mol. Microbiol. 45:1499-1513(2002).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=22222988; PubMed=12200547;
RA Tettelin H., Masiagnani V., Cieslewicz M.J., Eisen J.A., Peterson S.,
RA Messelt L.C., Paulsen I.T., Nelson K.E., Margalit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouli H., Mulligan S.,
RA Carthy H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Jacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Malone D.,
RA Rinaldo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae."
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
CC -1- FUNCTION: IF-3 binds to the 30S ribosomal subunit and shifts the
CC equilibrium between 70S ribosomes and their 50S and 30S subunits in
CC favor of the free subunits, thus enhancing the availability of 30S
CC subunits on which protein synthesis initiation begins.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the IF-3 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: AL766850; CAD47113.1; -
CC EMBL: AE014255; AAN00255.1; -
CC Sagalini: gbs1454; -
CC TIGR: SAG1384; -
CC HAMAP: MF_00080; -
CC Pfam: PF00707; IF3_C; 1.
CC Pfam: PF05198; IF3_N; 1.
CC ProDom: PD002880; IF3; 1.
CC TIGRFAMS: TIGR00168; infc; 1.
CC PROSITE: PS00938; IF3; 1.
CC Initiation factor: Protein biosynthesis; Complete proteome.
SQ SEQUENCE 176 AA; 20092 MW; 424B4EDDD814A2DA CRC64;

Query Match 1.7%; Score 7; DB 1; Length 176;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 112 OKEORRK 118

DB 77 OKEORRK 83
RESULT 10
ID IF3-STRA3 STANDARD: PRT: 176 AA.
AC O8DYT9:
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE Translation initiation factor IF-3.
GN INFC OR SMU.697.
OS Streptococcus mutans.
OC Bacteria: Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UA159 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397186;
RA Ajdic D., Mcshan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
RT pathogen."
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
CC -1- FUNCTION: IF-3 binds to the 30S ribosomal subunit and shifts the
CC equilibrium between 70S ribosomes and their 50S and 30S subunits in
CC favor of the free subunits, thus enhancing the availability of 30S
CC subunits on which protein synthesis initiation begins.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the IF-3 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: AE014913; AAN58429.1; -
CC HAMAP: MF_00080; -
CC Pfam: PF00707; IF3_C; 1.
CC Pfam: PF05198; IF3_N; 1.
CC ProDom: PD002880; IF3; 1.
CC TIGRFAMS: TIGR00168; infc; 1.
CC PROSITE: PS00938; IF3; 1.
CC Initiation factor: Protein biosynthesis; Complete proteome.
SQ SEQUENCE 176 AA; 20103 MW; A6873F87BFC3875 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 176;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 112 OKEORRK 118

DB 77 OKEORRK 83
RESULT 11
ID IF3-STRA3 STANDARD: PRT: 176 AA.
AC P58081;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Translation initiation factor IF-3.
GN INFC OR SPY0804 OR SPYM3_0538 OR SPS1316 OR SPYM18_0866.
OS Streptococcus pyogenes (serotype M3), and

OS Streptococcus pyogenes (serotype M18).
 CC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 CC Streptococcus.
 CC NCBI_TaxID=1314, 198466, 186103;
 CC [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
 RX MEDLINE=2192684; PubMed=11296296;
 RA Ferretti J.J., Mcshan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
 Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
 Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
 Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
 RT "Complete genome sequence of an M1 strain of Streptococcus
 pyogenes.";
 RT Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
 [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=MGAS315 / Serotype M3;
 RX MEDLINE=22133808; PubMed=12122206;
 RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
 Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Perkins L.D.,
 Campbell D.S., Smith T.M., McCormack J.K., Leung D.Y.M.,
 Schlievert P.M., Musser J.M.;
 RT "Genome sequence of a serotype M3 strain of group A Streptococcus:
 phage-encoded toxins, the high-virulence phenotype, and clone
 emergence.";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
 [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN=SSI-1 / Serotype M3;
 RA Nakagawa I., Kurokawa K., Nakata M., Tomiyasu Y., Yamashita A.,
 Yamazaki K., Okahashi N., Kawabata S., Yasunaga T., Hattori M.,
 Hayashi H., Hamada S.;
 RT "The genome of invasive Streptococcus pyogenes: a comparative analysis
 of S. pyogenes SSI-1, SF370 and MGAS8232.";
 RT Submitted (May-2002) to the EMBL/Genbank/DBJ databases.
 [4]
 RN SEQUENCE FROM N.A.
 RC STRAIN=MGAS8232 / Serotype M18;
 RX MEDLINE=21927593; PubMed=11917108;
 RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
 Sylva G.L., Studevant D.E., Ricklefs S.M., Porcella S.F.,
 Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
 Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
 RT "Genome sequence and comparative microarray analysis of serotype M18
 group A Streptococcus strains associated with acute rheumatic fever
 outbreaks.";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).
 CC -1- FUNCTION: IF-3 BINDS TO THE 30S RIBOSOMAL SUBUNIT AND SHIFTS THE
 EQUILIBRIUM BETWEEN 70S RIBOSOMES AND THEIR 50S AND 30S SUBUNITS IN
 FAVOR OF THE FREE SUBUNITS, THUS ENHANCING THE AVAILABILITY OF 30S
 SUBUNITS ON WHICH PROTEIN SYNTHESIS INITIATION BEGINS.
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to the IF-3 family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AEO06531; AAN33741.1;
 DR EMBL: AEO14147; AAM79145.1;
 DR EMBL: APO05145; BAC64411.1; ALT_INIT.
 DR EMBL: AEO10016; AAL97520.1;
 DR HSSP: P03000; I1TG.
 DR HAMAP: MF_00080; 1.
 DR InterPro: IPR001288; IF3.
 DR Pfam: PF00707; IF3_C; 1.
 DR Pfam: PF05198; IF3_N; 1.

DR ProDom: PD002880; IF3; 1.
 DR TIGRFAMS: TIGR00168; InfC; 1.
 DR PROSITE: PS00938; IF3; 1.
 KW Initiation factor; Protein biosynthesis; Complete proteome.
 SQ SEQUENCE 176 AA; 20053 MW; FD3247FB6C9D5919 CRC64;
 Query Match 1.7%; Score 7; DB 1; Length 176;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 112 OKEORRK 118
 Db 77 OKEORRK 83
 RESULT 12
 ID R4_HYMLI
 AC 020230;
 DT 15-DEC-1998 (rel. 37, Created)
 DT 15-DEC-1998 (rel. 37, Last sequence update)
 DT 28-FEB-2003 (rel. 41, Last annotation update)
 DE Chloroplast 30S ribosomal protein S4 (Fragment).
 GN RPS4.
 OS Hymenocallis littoralis (Beach spiderlily).
 OG Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Amaryllidaceae;
 OC Hymenocallis.
 CC NCBI_TaxID=59040;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Souza-Chies T.T., Bittar G., Nadot S., Carter L., Besin E.,
 Lejeune B.P.;
 RT "Phylogenetic analysis of Iridaceae with parsimony and distance
 RT methods using the plastid gene rps4.";
 RT Plant Syst. Evol. 204:109-123(1997).
 CC -1- FUNCTION: One of the primary rRNA binding proteins, it binds
 CC directly to 16S rRNA where it nucleates assembly of the body of
 CC the 30S subunit (By similarity).
 CC -1- FUNCTION: With S5 and S12 plays an important role in translational
 CC accuracy (By similarity).
 CC -1- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5.
 CC The interaction surface between S4 and S5 is involved in control
 CC of translational fidelity (By similarity).
 CC -1- SUBCELLULAR LOCATION: Chloroplast.
 CC -1- SIMILARITY: Contains 1 S4 RNA-binding domain.
 CC -1- SIMILARITY: BELONGS TO THE S4P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z68251; CA92549.1;
 DR HSSP: P81288; IC05.
 DR HAMAP: MF_01306; 1.
 DR InterPro: IPR001912; Ribosomal_S4.
 DR InterPro: IPR002942; S4.
 DR InterPro: IPR005709; S4_bact_org.
 DR Pfam: PF00163; Ribosomal_S4; 1.
 DR Pfam: PF01479; S4; 1.
 DR SMART: SM00363; S4; 1.
 DR TIGRFAMS: TIGR01017; rpsd_bact; 1.
 DR PROSITE: PS00632; RIBOSOMAL_S4; 1.
 DR PROSITE: PS50889; S4; 1.
 KW Ribosomal protein; RNA-binding; rRNA-binding; Chloroplast.
 KW NON_TER 1;
 FT DOMAIN 82 143 S4 RNA-BINDING.
 FT NON_TER 182 182

SQ SEQUENCE 182 AA; 21050 MW; 78B1261C3F6DC8AF CRC64;
 Query Match 1.7%; Score 7; DB 1; Length 182;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 VOIARAKA 64
 Db 62 VOIARAKA 68

RESULT 13
 IF3_STRPN STANDARD; PRT; 185 AA.
 AC Q97R70;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE Translation initiation factor IF-3.
 GN INFC OR SP0959 OR SP0861.
 OS Streptococcus pneumoniae, and
 OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_Taxid=1313, 171101.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC BAA-334 / TIGR4;
 RX MEDLINE=21357209; PubMed=11463916;
 RA Tettein H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson R.J.,
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radue D.,
 RA Holtzapple E., Khouli H., Wolf A.M., Uterback T.R., Hansen C.L.,
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickenson T., Hickey E.K.,
 RA Holt J.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.,
 RT "Complete genome sequence of a virulent isolate of Streptococcus
 RT pneumoniae."
 RL Science 293:498-506(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC BAA-255 / R6;
 RX MEDLINE=21492245; PubMed=11544234;
 RA Hoskins J., Aldorn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
 RA Dehoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Gerlinger C.,
 RA Leblanc D.J., Lee L.N., Leikowitz E.J., Lu J., Matsushima P.,
 RA McAhren S.M., McHenry M., Mcleaster K., Mundy C.W., Niclas T.I.,
 RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rokey P.,
 RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
 RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R. Jr., Skatrud P.L.,
 RT "Genome of the bacterium Streptococcus pneumoniae strain R6."
 RL J. Bacteriol. 183:5709-5717(2001).
 RN [3]
 RP FUNCTION: IF-3 binds to the 30S ribosomal subunit and shifts the
 RC equilibrium between 70S ribosomes and their 50S and 30S subunits in
 CC favor of the free subunits, thus enhancing the availability of 30S
 CC subunits on which protein synthesis initiation begins.
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to the IF-3 family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE007400; AAK75080.1;
 CC EMBL; AE008461; AAK99665.1; ALT_INIT.

DR PIR; G95110; G95110.
 DR TIGR; SP0959; -.
 DR HAMAP; MF_00080; -; 1.
 DR InterPro; IPR001288; IF3.
 DR Pfam; PF00707; IF3_C; 1.
 DR Pfam; PF05198; IF3_N; 1.
 DR ProDom; PD002880; IF3; 1.
 DR TIGRFAMs; TIGR00168; Infc; 1.
 DR PROSITE; PS00938; IF3; 1.
 KW Initiation factor; Protein biosynthesis; Complete proteome.
 SQ SEQUENCE 185 AA; 21165 MW; 422496FBB9463732 CRC64;

QY 112 OKEDRKK 118
 Db 86 OKEDRKK 92

Query Match 1.7%; Score 7; DB 1; Length 185;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 14
 RB24_HUMAN STANDARD; PRT; 203 AA.
 AC Q96905;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE Ras-related protein Rab-24.
 GN RAB24.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas, Skin, and uterus;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usgin T.B., Toshylyki S., Carninci P., Pringle C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Muliahy S.J.,
 RA Bosak S.A., McKen P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.M., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Scherren A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE OF 2-203 FROM N.A.
 RC TISSUE=Amygdala;
 RA Ottewill B., Obermaier B., Mewes H.-W., Gassenhuber J.,
 RA Wiemann S.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBP databases.
 CC -1- FUNCTION: MAY BE INVOLVED IN AUTOPHAGY-RELATED PROCESSES (By
 CC similarity).
 CC -1- SUBCELLULAR LOCATION: IN THE ENDOPLASMIC RETICULUM/CIS-GOLGI
 CC REGION AND ON LATE ENDOPLASMIC STRUCTURES (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RAB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; BC010006; AAH10006.1; -
DR EMBL; BC015534; AAH15534.1; -
DR EMBL; BC021263; AAH21263.1; -
DR EMBL; AL833898; CAD38754.1; -
DR Genew; HGNC:9765; RAB24.
DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR001806; Ras_trnsfmrng.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASRNSFRNG.
DR SMART; SM00175; Rab; 1.
DR TIGRfams; TIGR00231; small_GTP; 1.
KW GTP-binding; Lipoprotein; Prenylation; Protein transport; Autophagy.
FT NP_BIND 14 21 GTP (BY SIMILARITY).
FT NP_BIND 63 67 GTP (BY SIMILARITY).
FT NP_BIND 120 123 GTP (BY SIMILARITY).
FT DOMAIN 37 45 EFFECTOR REGION (BY SIMILARITY).
FT LIPID 200 200 GERANYL-GERANYL (BY SIMILARITY).
FT LIPID 201 201 GERANYL-GERANYL (BY SIMILARITY).
SQ SEQUENCE 203 AA; 23124 MW; 979E1AF6F7A4E5F1 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 203;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 ELSLSE 107
|111111|
DB 105 ELSLSE 111

RESULT 15
RB24_MOUSE STANDARD; PRT: 203 AA.
ID RB24_MOUSE
AC P35290.
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ras-related protein Rab-24 (Rab-16).
GN RAB24.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94171934; PubMed=8126105;
RA Oikonen V.M., Dupree P., Killisch I., Luetcke A., Simons K.,
RA Zerial M.;
RT "Molecular cloning and subcellular localization of three GTP-binding
RT proteins of the rab subfamily";
RL J. Cell Sci. 106:1249-1261(1993).
RN [2]
RP SEQUENCE OF 1-68 FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=92210010; PubMed=1555775;
RA Chavrier P., Simons K., Zerial M.;
RT "The complexity of the Rab and Rho GTP-binding protein subfamilies
RT revealed by a PCR cloning approach";
RL Gene 112:261-264(1992).
CC -1- FUNCTION: MAY BE INVOLVED IN AUTOPHAGY-RELATED PROCESSES.
CC -1- SUBCELLULAR LOCATION: IN THE ENDOPLASMIC RETICULUM/CIS-GOLGI
CC REGION AND ON LATE ENDOPLASMIC STRUCTURES.
CC -1- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RAB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; Z22819; CAA80472.1; -
DR EMBL; M79306; AAK14830.1; -
DR PIR; JH0646; JH0646.
DR PIR; S40235; S40235.
DR HSSP; P36017; IEK0.
DR MGD; MGI:105065; Rab24.
DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR001806; Ras_trnsfmrng.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR SMART; SM00175; Rab; 1.
DR TIGRfams; TIGR00231; small_GTP; 1.
KW GTP-binding; Lipoprotein; Prenylation; Protein transport; Autophagy.
FT NP_BIND 14 21 GTP (BY SIMILARITY).
FT NP_BIND 63 67 GTP (BY SIMILARITY).
FT NP_BIND 120 123 GTP (BY SIMILARITY).
FT DOMAIN 37 45 EFFECTOR REGION (BY SIMILARITY).
FT LIPID 200 200 GERANYL-GERANYL (BY SIMILARITY).
FT LIPID 201 201 GERANYL-GERANYL (BY SIMILARITY).
FT CONFLICT 67 67 S -> Q (IN REF. 2).
SQ SEQUENCE 203 AA; 23144 MW; 872EA5E2E7B5B0F0 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 203;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 ELSLSE 107
|111111|
DB 105 ELSLSE 111

Search completed: July 28, 2003, 09:28:00
Job time : 25 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2003, 09:24:47 ; Search time 99 Seconds
(without alignments)
1081.736 Million cell updates/sec

Title: US-09-068-377c-1

Perfect score: 415
Sequence: 1 MAAQLQFRDAFMCRDFTAH.....WTVERNQGRGVPGSYLEKL 415

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_23:*
1: sp._archaea:*
2: sp._bacteria:*
3: sp._fungi:*
4: sp._human:*
5: sp._invertebrate:*
6: sp._mammal:*
7: sp._mhc:*
8: sp._organelle:*
9: sp._phage:*
10: sp._plant:*
11: sp._rodent:*
12: sp._virus:*
13: sp._vertebrate:*
14: sp._unclassified:*
15: sp._viral:*
16: sp._bacteriophage:*
17: sp._archaeoprotein:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	415	100.0	415	11	P97814
2	31	7.5	397	4	O43585
3	31	7.5	416	4	O43586
4	31	7.5	416	4	O95657
5	10	2.4	4639	5	O9VZ83
6	8	1.9	106	10	O9SU61
7	8	1.9	332	12	O91ER7
8	8	1.9	344	10	O81E98
9	8	1.9	378	16	O8UF12
10	8	1.9	385	16	O8FN20
11	8	1.9	391	2	O93KB1
12	8	1.9	425	16	O8NT50
13	8	1.9	441	2	O813A5
14	8	1.9	457	5	O9VC07
15	8	1.9	513	16	O8G847
16	8	1.9	541	16	O69851

17	8	1.9	586	16	O9HZD4
18	8	1.9	629	16	O8DIZ3
19	8	1.9	633	3	O74307
20	8	1.9	856	5	O9N9R7
21	8	1.9	1147	16	O50466
22	8	1.9	1180	4	O92625
23	8	1.9	1218	5	O81ED2
24	8	1.9	1265	10	O8W5E1
25	7	1.7	63	16	O8XEM6
26	7	1.7	65	5	O8MNS7
27	7	1.7	77	17	O8Z2P2
28	7	1.7	82	7	O30604
29	7	1.7	88	2	O9ZET6
30	7	1.7	95	10	O9FM21
31	7	1.7	95	16	O9KX8
32	7	1.7	98	5	O27827
33	7	1.7	102	5	O8S218
34	7	1.7	105	17	O9YF08
35	7	1.7	111	12	O9WMC9
36	7	1.7	133	16	O8YJ10
37	7	1.7	133	16	O8EDJ4
38	7	1.7	134	2	O93P07
39	7	1.7	134	16	O811P9
40	7	1.7	135	2	O8KRY0
41	7	1.7	135	17	O97X99
42	7	1.7	136	5	O9VHNO
43	7	1.7	137	16	O8FYM9
44	7	1.7	140	10	O8LMD0
45	7	1.7	143	10	O94BB2

ALIGNMENTS

RESULT 1
P97814 PRELIMINARY: PRT: 415 AA.
ID P97814
AC P97814:
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE PEST phosphatase interacting protein.
GN PSTPIP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Lasky L.;
RT "PEST Phosphatase Interacting Protein (PSTPIP)";
RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL: U87814; AAB48483.1; -!
DR HSSP: Q60631; IGBO.
DR MGD: MGI:1321396; Pstpip1.
DR InterPro: IPR001060; Cdc15_Pes_CIP4.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00611; FCH; 1.
DR Pfam: PF00018; SH3; 1.
DR PRINTS: PRO0452; SH3DOMAIN.
DR PRODOM: PD000066; SH3; 1.
DR SMART: SM00326; FCH; 1.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS50133; FCH; 1.
DR PROSITE: PS50002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 415 AA: 47590 MW: 1600329284D2739C CRC64:
Query Match 100.0%; Score 415; DB 11; Length 415;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MAAOLQFRAFWCQDFTAHGTEVLLQRLDGRKMKCDVEELLQRAQAEEYRGKELVQI 60
DB 1 MAAOLQFRAFWCQDFTAHGTEVLLQRLDGRKMKCDVEELLQRAQAEEYRGKELVQI 60
QY 61 ARKAGGOTEMNSLTSTSDSLKOOTENYSAHIQIALALRELSLEFFERQKQKYE 120
DB 61 ARKAGGOTEMNSLTSTSDSLKOOTENYSAHIQIALALRELSLEFFERQKQKYE 120
QY 121 AIMDRVQKSLSLYKKTMTESKKAAYDQKCRDADAEQAERFVSANGHKOYKESONKAKOC 180
DB 121 AIMDRVQKSLSLYKKTMTESKKAAYDQKCRDADAEQAERFVSANGHKOYKESONKAKOC 180
QY 181 KESATAEARYRONIEOLERARTEWEDEHRTTCEAFOLQEFDRLLTLRNALMWHCQNLQM 240
DB 181 KESATAEARYRONIEOLERARTEWEDEHRTTCEAFOLQEFDRLLTLRNALMWHCQNLQM 240
QY 241 QCVKDELYEVRLLTEGCGVEGJINGEIOKSGRPPAPVPYONYNDEYVPLIGSPS 300
DB 241 QCVKDELYEVRLLTEGCGVEGJINGEIOKSGRPPAPVPYONYNDEYVPLIGSPS 300
QY 301 IQPSGCVIKRFSGLHSGPPTPSAPASTETLTPPERNELVYASIEVQATQGNLNSSA 360
DB 301 IQPSGCVIKRFSGLHSGPPTPSAPASTETLTPPERNELVYASIEVQATQGNLNSSA 360
QY 361 QDPRALYDYTAQNSDELDISAGDILAVILEGEGMWTVERNNGRGVPGSYLEKL 415
DB 361 QDPRALYDYTAQNSDELDISAGDILAVILEGEGMWTVERNNGRGVPGSYLEKL 415

```

RESULT 2

```

ID 043585 PRELIMINARY; PRT; 397 AA.
AC 043585;
DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE CD2 binding protein 1 short form.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN 1;
RP SEQUENCE FROM N.A.
RX MEDLINE=99077800; PubMed=9857189;
RA Li J., Nishizawa K., An W., Hussey R.E., Lialios F.E., Saigia R.,
RA Sunder-Plassmann R., Reinherz E.L.;
RT "A cdcl5-like adaptor protein (CD2BP1) interacts with the CD2
RT cytoplasmic domain and regulates CD2-triggered adhesion.";
RL EMBL J. 17:7320-7336(1998).
CC -1- SIMILARTY: CONTAINS 1 SH3 DOMAIN.
DR EMBL: AF038603; AAD11959.1; -.
DR HSSP: P07751; 1BK2.
DR InterPro: IPR001060; Cdc15_Fes_CIP4.
DR pfam: PF00611; FCH; 1.
DR PRINTS: PR00452; SH3DOMAIN.
DR ProDom: PD000066; SH3; 1.
DR SMART: SM00055; FCH; 1.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS50133; FCH; 1.
DR PROSITE: PS50002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 397 AA: 45353 MW: 54D64A3AAE16A2FC CRC64;

```

Query Match 7.5%; Score 31; DB 4; Length 397;
 Best Local Similarity 100.0%; Pred. No. 3.2e-23;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 204 EWEQEHRTTCEAFQLOEFDRLTLRNALMVH 234
DB 204 EWEQEHRTTCEAFQLOEFDRLTLRNALMVH 234

```

RESULT 3

```

ID 043586 PRELIMINARY; PRT; 416 AA.
AC 043586;
DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE CD2 binding protein 1 long form (Proline-serine-threonine phosphatase
DE interacting protein 1).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN 1;
RP SEQUENCE FROM N.A.
RX MEDLINE=99077800; PubMed=9857189;
RA Li J., Nishizawa K., An W., Hussey R.E., Lialios F.E., Saigia R.,
RA Sunder-Plassmann R., Reinherz E.L.;
RT "A cdcl5-like adaptor protein (CD2BP1) interacts with the CD2
RT cytoplasmic domain and regulates CD2-triggered adhesion.";
RL EMBL J. 17:7320-7336(1998).
CC -1- SIMILARTY: CONTAINS 1 SH3 DOMAIN.
DR EMBL: AF038603; AAD11959.1; -.
DR EMBL: BC008602; AAO08602.1; -.
DR HSSP: P07751; 1BK2.
DR Genew: HGNC:9580; PSTPIP1.
DR InterPro: IPR001060; Cdc15_Fes_CIP4.
DR pfam: PF00611; FCH; 1.
DR PRINTS: PR00452; SH3DOMAIN.
DR ProDom: PD000066; SH3; 1.
DR SMART: SM00055; FCH; 1.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS50133; FCH; 1.
DR PROSITE: PS50002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 416 AA: 47591 MW: 97818150B3D5D600 CRC64;

```

Query Match 7.5%; Score 31; DB 4; Length 416;
 Best Local Similarity 100.0%; Pred. No. 3.3e-23;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 204 EWEQEHRTTCEAFQLOEFDRLTLRNALMVH 234
DB 204 EWEQEHRTTCEAFQLOEFDRLTLRNALMVH 234

```

RESULT 4

```

ID 095657 PRELIMINARY; PRT; 416 AA.
AC 095657;
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE PEST phosphatase interacting protein homolog.
GN H-PIP.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN 1;
RP SEQUENCE FROM N.A.
RX Wilson L.A., Fields D., Cruz L., Lasky L., Friesen J.,
RX Siminovitch K.A.;
RT "The human homologue of mouse PTP-PIP interactor protein.";
RT Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.

```

```
CC -i SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL: U94778; AAD00762.1; -.
DR HSSP: 060631; 1GB0.
DR InterPro: IPR001060; Cdc15_Fes_C1P4.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00611; FCH; 1.
DR Pfam: PF00018; SH3; 1.
DR PRINTS: PR00452; SH3DOMAIN.
DR ProDom: PD000066; SH3; 1.
DR SMART: SM00055; FCH; 1.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS50133; FCH; 1.
DR PROSITE: PS50002; SH3; 1.
DR SH3 domain.
SQ SEQUENCE 416 AA; 47610 MW; 7CED2A42E1A7EEFF CRC64;

Query Match 7.5%; Score 31; DB 4; Length 416;
Best Local Similarity 100.0%; Pred. No. 3.3e-23;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 204 EMEQHRRTCEAFQLOEFDRLTLRNALMVH 234
Db 204 EMEQHRRTCEAFQLOEFDRLTLRNALMVH 234

RESULT 5
O9V283 PRELIMINARY: PRT; 4639 AA.
AC O9V283:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE CG7507 protein.
ID DHC64C OR CG7507.
GN Drosophila melanogaster (Fruit fly).
OS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernick B.P., Bhandari D., Bolshakov S.,
RA Borova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glisak A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Sidenkiam I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Sylrskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weisslock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).

[2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galie R.F., Gary N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattel B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunco J.,
RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phoumenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Sylrskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

[3]
RP SEQUENCE FROM N.A.
RA Mistra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kamlirker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Bertram B., Carlson J.W., Celniker S.E.,
RA Clamp M., Dysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutnak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

[4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

[5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AEO03482; AAF7942.3; -.
DR FlyBase: FBgn0010349; Dhc64C.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR001064; Crystallin.
DR InterPro: IPR004273; Dynein_heavy.
DR InterPro: IPR00169; SHprot_acsile.
DR Pfam: PF03028; Dynein_heavy; 1.
DR SMART: SM00382; AAA; 4.
DR PROSITE: PS00225; CRYSTALLIN_BETRAGAMA; 1.
DR PROSITE: PS00639; THIOL_PROTEASE_HIS; 1.
KW ATP-binding.
SQ SEQUENCE 4639 AA; 530172 MW; 2A3DAF1FBBE52096 CRC64;

Query Match 2.4%; Score 10; DB 5; Length 4639;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 98 LREELNSLEE 107
Db 3398 LREELNSLEE 3407

RESULT 6
O9S061 PRELIMINARY: PRT; 106 AA.
ID O9S061:
AC O9S061:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
```

DE Hypothetical 11.1 kda protein.
 GN T17P15.150.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Eumetazoa; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eusterozoa; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Choisme N., Robert C., Brottier P., Wincker P., Cattolico L.,
 RA Artiguenave F., Saurin W., Weissenbach J., Mewes H.W., Mayer K.F.X.,
 RA Lemcke K., Schueller C., Queller F., Salanoubat M.,
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project.
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL049658; CAB41141.1; -
 DR InterPro; IPR007065; HPP; 1.
 DR Pfam; PF04982; HPP; 1.
 KW Hypothetical protein.
 SO SEQUENCE 106 AA; 11094 MW; ECD19CDB5C191C3 CRC64;

Query Match 1.9%; Score 8; DB 10; Length 106;
 Best Local Similarity 100.0%; Pred. No. 8.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 321 TTSPAPAA 328
 Db 32 TTSPAPAA 39

RESULT 7
 ID 091ER7 PRELIMINARY; PRT; 332 AA.
 AC 091ER7;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE ORF138 VP1054.
 GN ORF138 VP1054.
 OS Cydia pomonella granulosis virus (CPGV) (Cydia pomonella
 OS granulovirus).
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae; Granulovirus.
 OX NCBI_TaxID=28289;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=Mexican 1;
 RX MEDLINE=9318168; PubMed=8445726;
 RA Crook N.E., Clem R.J., Miller L.K.,
 RT "An apoptosis-inhibiting baculovirus gene with a zinc finger-like
 RT motif."
 RL J. Virol. 67:2168-2174(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA STRAIN=Mexican 1;
 RX MEDLINE=96207404; PubMed=8615018;
 RA Theilmann D.A., Chantler J.K., Stewart S., Flipsen H.T., Vlak J.M.,
 RA Crook N.E.,
 RT "Characterization of a highly conserved baculovirus structural protein
 RT that is specific for occlusion-derived virions."
 RL Virology 218:148-158(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA STRAIN=Mexican 1;
 RX MEDLINE=97380577; PubMed=9237352;
 RA Kang W., Crook N.E., Winstanley D., O'Reilly D.R.,
 RT "Complete sequence and transposon mutagenesis of the BamHI J fragment
 RT of Cydia pomonella granulosis virus."
 RL Virus genes 14:131-136(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA STRAIN=Mexican 1;

EX MEDLINE=96418511; PubMed=9747739;
 RA Kang W., Tristem M., Maeda S., Crook N.E., O'Reilly D.R.,
 RT "Identification and characterization of the Cydia pomonella
 RT granulovirus cathepsin and chitinase genes."
 RL J. Gen. Virol. 79:2283-2292(1998).
 RN [5]
 RP SEQUENCE FROM N.A.
 RA STRAIN=Mexican 1;
 RC Luque T., Finch R., Crook N., O'Reilly D.R., Winstanley D.,
 RA "The complete sequence of the Cydia pomonella granulovirus genome."
 RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U53466; AKR70798.1; -
 SO SEQUENCE 332 AA; 38765 MW; E4C20584D3063482 CRC64;

Query Match 1.9%; Score 8; DB 12; Length 332;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 EELRSLEE 107
 Db 98 EELRSLEE 105

RESULT 8
 ID 08LE98 PRELIMINARY; PRT; 344 AA.
 AC 08LE98;
 DT 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Hypothetical protein.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eusterozoa; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Haas B.J., Volkovsky N., Town C.D., Troukhan M., Alexandrov N.,
 RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.,
 RT "Full-length messenger RNA sequences greatly improve genome
 RT annotation."
 RL Genome Biol. 0:0-0(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
 RA Feldmann K.,
 RT "Full-length cDNA from Arabidopsis thaliana."
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY085543; AAM62767.1; -
 KW Hypothetical protein
 SO SEQUENCE 344 AA; 38357 MW; 84B92E14C9FB2861 CRC64;

Query Match 1.9%; Score 8; DB 10; Length 344;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 ESATEAER 189
 Db 117 ESATEAER 124

RESULT 9
 ID 08UF12 PRELIMINARY; PRT; 378 AA.
 AC 08UF12;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Inosine-uridine preferring nucleoside hydrolase.
 GN IUNH OR ATU1590 OR AGR-C-2923.
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OY 322 TPSAPAS 329
 |||||||
 DB 393 TPSAPAS 400

RESULT 13

ID 0813A5 PRELIMINARY; PRT: 441 AA.
 AC 0813A5;
 DT 01-OCT-2002 (TREMblrel. 22, Created)
 DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)
 DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)
 DE BSS1 restriction endonuclease.
 GN BSSSR.
 OS Bacillus stearothermophilus.
 OC Bacteria; Firmicutes; Bacillales; Geobacillus.
 OX NCBI_TaxId=1422;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5719;
 RT Helter D., Wilson G.G.;
 RL "Sequence of the BSS1 restriction-modification genes."
 DR Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
 KW EMBL: AF254788; AAM21166.1; -
 SO SEQUENCE 441 AA; 51784 MW; 458B34DC8EE21050 CRC64;

Query Match 1.9%; Score 8; DB 2; Length 441;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 268 FIOSKSTG 275
 |||||||
 DB 312 FIOSKSTG 319

RESULT 14

ID 09VGU7 PRELIMINARY; PRT: 457 AA.
 AC 09VGU7;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
 DE CG14696 protein.
 GN CG14696.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxId=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amaratilake P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe C.R., Pfeiffer B.D.,
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agayuni A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris J.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Ventler E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.W., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003689; AAF54577.1; -
 DR FlyBase: FBgn0037853; CG14696.
 SO SEQUENCE 457 AA; 50052 MW; C1CBDACBC20E3BA3 CRC64;

Query Match 1.9%; Score 8; DB 5; Length 457;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 322 TPSAPAS 329
 |||||||
 DB 442 TPSAPAS 449

RESULT 15

ID 08G847 PRELIMINARY; PRT: 513 AA.
 AC 08G847;
 DT 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE ATP binding protein of ABC transporter.
 GN BFO034.
 OS Bifidobacterium longum.
 OC Bacteria; Actinobacteria; Actinobacteriales; Bifidobacteriales;
 OC Bifidobacteriaceae; Bifidobacterium.
 OX NCBI_TaxId=216816;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NC-2705;
 RX MEDLINE=22294977; PubMed=12381787;
 RA Schell M.A., Karimiantzou M., Snel B., Vilanova D., Berger B.,
 RA Pessi G., Zwaalen M.-C., Desiere F., Bork P., Delley M.,
 RA Pridmore R.D., Arlioni F.;
 RT "The genome sequence of Bifidobacterium longum reflects its adaptation
 to the human gastrointestinal tract."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
 DR EMBL: AE014617; AMN23901.1; -
 KW Complete proteome.
 SO SEQUENCE 513 AA; 55835 MW; F47E2AFB6798AB9 CRC64;

Query Match 1.9%; Score 8; DB 16; Length 513;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 222 DRILLILRN 229
 |||||||
 DB 212 DRILLILRN 219

Search completed: July 28, 2003, 09:29:48
 Job time: 103 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2003, 09:26:02 ; Search time 29 Seconds
(without alignments)
605.483 Million cell updates/sec.

Title: US-09-068-377C-1

Perfect score: 415
Sequence: 1 MMAQLQFRDAFWCRRDFTAHGTVLLQRLIDGRKMKDVEELLRQRAQAEERYGKEIVQI 415

Scoring table: OLIGO
Gapop 60.0, Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size: 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: Issued_Patents_AA:*

1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/FCITUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	415	100.0	415	US-08-938-830-1	Sequence 1, Appl
2	415	100.0	415	US-09-020-222-1	Sequence 1, Appl
3	31	7.5	166	US-09-006-428A-3	Sequence 3, Appl
4	31	7.5	397	US-09-006-428A-2	Sequence 2, Appl
5	31	7.5	397	US-09-006-428A-19	Sequence 19, Appl
6	31	7.5	400	US-08-938-830-29	Sequence 29, Appl
7	31	7.5	416	US-09-006-428A-1	Sequence 1, Appl
8	31	7.5	416	US-09-006-428A-17	Sequence 17, Appl
9	29	7.0	57	US-09-006-428A-5	Sequence 5, Appl
10	24	5.8	48	US-08-938-830-3	Sequence 3, Appl
11	24	5.8	48	US-09-020-222-3	Sequence 3, Appl
12	8	1.9	923	US-09-252-991A-28964	Sequence 28964, A
13	7	1.7	57	US-08-630-915A-136	Sequence 136, App
14	7	1.7	60	US-08-630-915A-112	Sequence 112, App
15	7	1.7	79	US-09-006-428A-14	Sequence 14, Appl
16	7	1.7	144	US-08-630-915A-6	Sequence 6, Appl
17	7	1.7	144	US-09-252-991A-20574	Sequence 20574, A
18	7	1.7	186	US-09-252-991A-30079	Sequence 30079, A
19	7	1.7	203	US-08-916-901-1	Sequence 1, Appl
20	7	1.7	203	US-08-916-901-7	Sequence 1, Appl
21	7	1.7	203	US-09-154-602-1	Sequence 1, Appl
22	7	1.7	203	US-09-154-602-7	Sequence 7, Appl
23	7	1.7	203	US-09-252-991A-23165	Sequence 23165, A
24	7	1.7	282	US-09-252-991A-20666	Sequence 20666, A
25	7	1.7	290	US-09-252-991A-24770	Sequence 24770, A
26	7	1.7	356	US-09-252-991A-18206	Sequence 18206, A
27	7	1.7	392	US-08-311-731A-210	Sequence 210, App

28	7	1.7	399	4	US-09-252-991A-22853	Sequence 22853, A
29	7	1.7	399	4	US-09-328-352-8043	Sequence 8043, Ap
30	7	1.7	411	4	US-09-134-001C-3121	Sequence 3121, Ap
31	7	1.7	462	4	US-08-630-915A-38	Sequence 38, Appl
32	7	1.7	508	3	US-09-111-730-2	Sequence 2, Appl
33	7	1.7	516	4	US-09-252-991A-28894	Sequence 28894, A
34	7	1.7	525	1	US-08-077-939-19	Sequence 19, Appl
35	7	1.7	525	1	US-08-461-599-19	Sequence 19, Appl
36	7	1.7	525	1	US-08-461-621-19	Sequence 19, Appl
37	7	1.7	525	1	US-08-465-334-19	Sequence 19, Appl
38	7	1.7	528	4	US-09-328-352-6385	Sequence 6385, Ap
39	7	1.7	573	1	US-08-200-512-2	Sequence 2, Appl
40	7	1.7	608	4	US-09-489-847-213	Sequence 213, App
41	7	1.7	664	4	US-09-328-352-6193	Sequence 6193, Ap
42	7	1.7	688	4	US-09-252-991A-17969	Sequence 17969, A
43	7	1.7	701	4	US-09-252-991A-26940	Sequence 26940, A
44	7	1.7	716	4	US-09-489-847-370	Sequence 370, App
45	7	1.7	722	3	US-08-390-874C-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-08-938-830-1
; Sequence 1, Application US/08938830
; Patent No. 6040437
; GENERAL INFORMATION:
; APPLICANT: Lasky, Laurence A.
; APPLICANT: Dowbenko, Donald J.
; TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage
; TITLE OF INVENTION: Furrow-Associated Proteins (PSRIPs)
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 Inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Minipalin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938, 830
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/798419
; FILING DATE: 07-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: P1066P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-3216
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 415 amino acids
; TYPE: Amino acid
; TOPOLOGY: Linear
; US-08-938-830-1
Query Match 100.0%; Score 415; DB 3; Length 415;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MMAQLQFRDAFWCRRDFTAHGTVLLQRLIDGRKMKDVEELLRQRAQAEERYGKEIVQI 60
DB 1 MMAQLQFRDAFWCRRDFTAHGTVLLQRLIDGRKMKDVEELLRQRAQAEERYGKEIVQI 60


```

; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-006-428A-2

```

```

Query Match
Best Local Similarity 100.0%; Score 31; DB 4; Length 397;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 204 EMEQEHRTCEAFQLOEFDRLLTLRNALMVH 234
|||||
Db 204 EMEQEHRTCEAFQLOEFDRLLTLRNALMVH 234

```

```

RESULT 5
US-09-006-428A-19
; Sequence 19, Application US/09006428A
; Patent No. 644439
; GENERAL INFORMATION:
; APPLICANT: Jing Li
; APPLICANT: Kazuhisa Nishizawa
; APPLICANT: Wenqian An
; APPLICANT: Ellis L. Reinherz
; TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF A
; FILE REFERENCE: 1062.1020-000
; CURRENT APPLICATION NUMBER: US/09/006,428A
; CURRENT FILING DATE: 1998-01-13
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-006-428A-19

```

```

Query Match
Best Local Similarity 100.0%; Score 31; DB 4; Length 397;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 204 EMEQEHRTCEAFQLOEFDRLLTLRNALMVH 234
|||||
Db 204 EMEQEHRTCEAFQLOEFDRLLTLRNALMVH 234

```

```

RESULT 6
US-08-938-830-29
; Sequence 29, Application US/08938830
; Patent No. 6040437
; GENERAL INFORMATION:
; APPLICANT: Lasky, Laurence A.
; APPLICANT: Dowbenko, Donald J.
; TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage
; TITLE OF INVENTION: Furrow-Associated Proteins (PSPPIPs)
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Winpatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,830
; FILING DATE:

```

```

; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/798419
; FILING DATE: 07-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: P1066P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-3216
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 400 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
US-08-938-830-29

```

```

Query Match
Best Local Similarity 100.0%; Score 31; DB 3; Length 400;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 204 EMEQEHRTCEAFQLOEFDRLLTLRNALMVH 234
|||||
Db 223 EMEQEHRTCEAFQLOEFDRLLTLRNALMVH 253

```

```

RESULT 7
US-09-006-428A-1
; Sequence 1, Application US/09006428A
; Patent No. 644439
; GENERAL INFORMATION:
; APPLICANT: Jing Li
; APPLICANT: Kazuhisa Nishizawa
; APPLICANT: Wenqian An
; APPLICANT: Ellis L. Reinherz
; TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF A
; FILE REFERENCE: 1062.1020-000
; CURRENT APPLICATION NUMBER: US/09/006,428A
; CURRENT FILING DATE: 1998-01-13
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-006-428A-1

```

```

Query Match
Best Local Similarity 100.0%; Score 31; DB 4; Length 416;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 204 EMEQEHRTCEAFQLOEFDRLLTLRNALMVH 234
|||||
Db 204 EMEQEHRTCEAFQLOEFDRLLTLRNALMVH 234

```

```

RESULT 8
US-09-006-428A-17
; Sequence 17, Application US/09006428A
; Patent No. 644439
; GENERAL INFORMATION:
; APPLICANT: Jing Li
; APPLICANT: Kazuhisa Nishizawa
; APPLICANT: Wenqian An
; APPLICANT: Ellis L. Reinherz
; TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF A
; FILE REFERENCE: 1062.1020-000
; CURRENT APPLICATION NUMBER: US/09/006,428A
; CURRENT FILING DATE: 1998-01-13
; NUMBER OF SEQ ID NOS: 28

```

SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 17
LENGTH: 416
TYPE: PRT
ORGANISM: Homo sapiens
US-09-006-428A-17

Query Match
Best Local Similarity 100.0%; Score 31; DB 4; Length 416;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 EWEDEHRTCEAFQLOEFDRITLRLNALVNH 234
DB 204 EWEDEHRTCEAFQLOEFDRITLRLNALVNH 234

RESULT 9
US-09-006-428A-5
Sequence 5, Application US/09006428A
Patent No. 6444439
GENERAL INFORMATION:
APPLICANT: Jimg Li
APPLICANT: Kazuhisa Nishizawa
APPLICANT: Menglan An
APPLICANT: Ellis L. Reinherz

TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF A
FILE REFERENCE: 1062.1020-000
CURRENT APPLICATION NUMBER: US/09/006,428A
CURRENT FILING DATE: 1998-01-13
NUMBER OF SEQ ID NOS: 28
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 5
LENGTH: 57
TYPE: PRT
ORGANISM: Homo sapiens
US-09-006-428A-5

Query Match
Best Local Similarity 100.0%; Score 29; DB 4; Length 57;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 387 VILEGEDGMWTVERNRGQFVPGSYLEKL 415
DB 29 VILEGEDGMWTVERNRGQFVPGSYLEKL 57

RESULT 10
US-08-938-830-3
Sequence 3, Application US/08938830
Patent No. 6040437
GENERAL INFORMATION:

APPLICANT: Lasky, Laurence A.
APPLICANT: Dowbenko, Donald J.
TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage
TITLE OF INVENTION: Furrow-Associated Proteins (FSTRIPS)
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938,830
FILING DATE:
CLASSIFICATION: 530

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/798419
FILING DATE: 07-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: P1066P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-3216
TELEFAX: 650/952-9881

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 48 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-938-830-3

Query Match
Best Local Similarity 100.0%; Score 24; DB 3; Length 48;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 366 LYDTRQNSDELDISAGDILAVIL 389
DB 1 LYDTRQNSDELDISAGDILAVIL 24

RESULT 11
US-09-020-222-3
Sequence 3, Application US/09020222
Patent No. 611073
GENERAL INFORMATION:

APPLICANT: Lasky, Laurence A.
TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage
TITLE OF INVENTION: Furrow-Associated Proteins (FSTRIPS)
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/020,222
FILING DATE: 06-Feb-1998
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/798419
FILING DATE: 02/07/1997
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: P1066-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-3216
TELEFAX: 650/952-9881

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 48 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-09-020-222-3

Query Match
Best Local Similarity 100.0%; Score 24; DB 3; Length 48;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 366 LYDTRQNSDELDISAGDILAVIL 389

Db 1 LYDYTAONSDLEISAGDILAVIL 24

RESULT 12

US-09-252-991A-28964
Sequence 28964, Application US/09232991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 28964
LENGTH: 923
TYPE: PRP
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28964

Query Match 1.9%; Score 8; DB 4; Length 923;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 39 VELLROR 46
11111111
Db 789 VELLROR 796

RESULT 13

US-08-630-915A-136
Sequence 136, Application US/08630915A
Patent No. 6309820
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
APPLICANT: HOFFMAN, No. 6309820h
APPLICANT: KAY, Brian K.
APPLICANT: FOWLES, Dana M.
APPLICANT: MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8664/9741
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 136:
SEQUENCE CHARACTERISTICS:
LENGTH: 57 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-630-915A-136

Query Match 1.7%; Score 7; DB 4; Length 57;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 367 YDYTAON 373
11111111
Db 9 YDYTAON 15

RESULT 14

US-08-630-915A-112
Sequence 112, Application US/08630915A
Patent No. 6309820
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
APPLICANT: HOFFMAN, No. 6309820h
APPLICANT: KAY, Brian K.
APPLICANT: FOWLES, Dana M.
APPLICANT: MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8664/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 112:
SEQUENCE CHARACTERISTICS:
LENGTH: 60 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-630-915A-112

Query Match 1.7%; Score 7; DB 4; Length 60;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 371 AONSDLE 377
11111111
Db 13 AONSDLE 19

RESULT 15
 US-09-006-428A-14
 ; Sequence 14, Application US/09006428A
 ; Patent No. 6444439
 ; GENERAL INFORMATION:
 ; APPLICANT: Jing Li
 ; APPLICANT: Kazuhisa Nishizawa
 ; APPLICANT: Wenqian An
 ; APPLICANT: Ellis L. Reinherz
 ; TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF A
 ; FILE REFERENCE: 1062.1020-000
 ; CURRENT APPLICATION NUMBER: US/09/006,428A
 ; CURRENT FILING DATE: 1998-01-13
 ; NUMBER OF SEQ ID NOS: 28
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 14
 ; LENGTH: 79
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-006-428A-14

Query Match 1.7%; Score 7; DB 4; Length 79;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 7; Conservative 0; Mismatches 0; Indels 0;

OY 363 YRALYDY 369
 |||||
 Db 8 YRALYDY 14

Search completed: July 28, 2003, 09:31:10
 Job time : 30 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2003, 09:29:52 ; Search time 52 Seconds
(without alignments)
947.794 Million cell updates/sec

Title: US-09-068-377c-1

Perfect score: 415
Sequence: 1 MAAQLQFRDAFWCRDFTAH.....WTERNGRGVPGSYLEKL 415

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 451899 seqs, 118759770 residues

Word size : 0

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: Published_Applications_AA.*
2: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/PCCT_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09C_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10C_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	195	47.0	323	15	US-10-067-076-2
2	92	22.2	92	15	US-10-067-076-3
3	31	7.5	416	15	US-10-067-076-1
4	31	7.5	416	15	US-10-067-076-20
5	30	7.2	416	15	US-10-067-076-22
6	8	1.9	425	15	US-09-738-626-4016
7	8	1.9	509	15	US-10-142-231-67
8	8	1.9	1139	16	US-10-080-170-460
9	7	1.7	57	9	US-09-879-957-136
10	7	1.7	60	9	US-09-879-957-112
11	7	1.7	70	9	US-09-864-761-42794
12	7	1.7	111	10	US-09-764-868-1002
13	7	1.7	144	9	US-09-879-957-6
14	7	1.7	155	10	US-09-860-670-87
15	7	1.7	186	9	US-09-764-853-550

16	7	1.7	195	9	US-09-815-242-13301	Sequence 13301, A
17	7	1.7	203	10	US-09-967-736-1	Sequence 1, Appl1
18	7	1.7	203	10	US-09-967-736-7	Sequence 7, Appl1
19	7	1.7	208	10	US-09-764-868-1104	Sequence 1104, Ap
20	7	1.7	210	9	US-09-800-729-181	Sequence 181, App
21	7	1.7	219	15	US-10-156-761-7982	Sequence 7982, Ap
22	7	1.7	223	10	US-09-764-868-681	Sequence 681, App
23	7	1.7	263	11	US-09-769-787-119	Sequence 119, App
24	7	1.7	287	15	US-10-156-761-12443	Sequence 12443, A
25	7	1.7	340	15	US-10-156-761-14143	Sequence 14143, A
26	7	1.7	352	11	US-09-820-843A-23	Sequence 23, Appl
27	7	1.7	361	15	US-09-815-242-10232	Sequence 10232, A
28	7	1.7	381	15	US-10-156-761-13257	Sequence 13257, A
29	7	1.7	390	9	US-09-815-242-5284	Sequence 5284, Ap
30	7	1.7	381	10	US-09-908-931B-22	Sequence 22, Appl
31	7	1.7	398	9	US-09-815-242-12599	Sequence 12599, A
32	7	1.7	404	15	US-10-106-698-4805	Sequence 4805, Ap
33	7	1.7	436	10	US-09-712-363-258	Sequence 258, App
34	7	1.7	443	9	US-09-815-242-5226	Sequence 5226, Ap
35	7	1.7	443	9	US-09-815-242-12520	Sequence 12520, A
36	7	1.7	454	10	US-09-771-161A-95	Sequence 95, Appl
37	7	1.7	462	9	US-09-879-957-38	Sequence 38, Appl
38	7	1.7	492	9	US-09-764-898-206	Sequence 206, App
39	7	1.7	495	15	US-10-156-761-10045	Sequence 10045, A
40	7	1.7	505	10	US-09-771-161A-186	Sequence 186, App
41	7	1.7	506	15	US-10-156-761-12648	Sequence 12648, A
42	7	1.7	508	9	US-09-878-262B-2	Sequence 2, Appl1
43	7	1.7	539	9	US-09-815-242-10497	Sequence 10497, A
44	7	1.7	544	14	US-10-047-542-73	Sequence 73, Appl
45	7	1.7	551	9	US-09-818-143-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-10-067-076-2
; Sequence 2, Application US/10067076
; Publication NO. US20030104404A1
; GENERAL INFORMATION:
; APPLICANT: WISE, Carol A.
; TITLE OF INVENTION: Genetic Markers for Autoimmune Disorder
; FILE REFERENCE: TEX871/4-006US/36000
; CURRENT APPLICATION NUMBER: US/10/067, 076
; CURRENT FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: 60/287, 893
; PRIOR FILING DATE: 2001-05-01
; PRIOR APPLICATION NUMBER: 09/710, 693
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Murine Amino Acid
US-10-067-076-2

Query Match 47.0%; Score 195; DB 15; Length 323;
Best Local Similarity 99.7%; Pred. No. 3, 1e-181;
Matches 295; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
DB 148 CRDADDAEAFERVSANGHOKOEVKSQNKAKCKESATAEARYYRONIQLEFARTEMQ 207
148 CRDADDAEAFERVSANGHOKOEVKSQNKAKCKESATAEARYYRONIQLEFARTEMQ 207

QY 208 EHRTCEAFQLOEFDRLTILRNALWVHNCQISMOCVDELYEYVRLTLESGDVEGDNG 267
DB 208 EHRTCEAFQLOEFDRLTILRNALWVHNCQISMOCVDELYEYVRLTLESGDVEGDNG 267
QY 268 FIOSKSTREPPAPYQNYDYREVTPLIGSPSIQPSGCVIKRFGSLHGSFKTTP 323
DB 268 FIOSKSTREPPAPYQNYDYREVTPLIGSPSIQPSGCVIKRFGSLHGSFKTTP 323

RESULT 2
US-10-067-076-3
; Sequence 3, Application US/10067076
; Publication No. US20030104404A1
; GENERAL INFORMATION:
; APPLICANT: Wise, Carol A.
; TITLE OF INVENTION: Genetic Markers for Autoimmune Disorder
; FILE REFERENCE: TEX871/4-006US/36000
; CURRENT APPLICATION NUMBER: US/10/067,076
; CURRENT FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: 60/287,893
; PRIOR FILING DATE: 2001-05-01
; PRIOR APPLICATION NUMBER: 09/710,693
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Murine Amino Acid
US-10-067-076-3

Query Match
Best Local Similarity 100.0%; Pred. No. 1.4e-81;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 324 SAPAESTELTPTPERNELVYASIEVQATQGNLNSSAQDYRALLYDTAQNDELDISAGD 383
DB 1 SAPAESTELTPTPERNELVYASIEVQATQGNLNSSAQDYRALLYDTAQNDELDISAGD 60
QY 384 ILAVILEGSDGWMYVRNGCRGVPGSYLEKL 415
DB 61 ILAVILEGSDGWMYVRNGCRGVPGSYLEKL 92

RESULT 3
US-10-067-076-1
; Sequence 1, Application US/10067076
; Publication No. US20030104404A1
; GENERAL INFORMATION:
; APPLICANT: Wise, Carol A.
; TITLE OF INVENTION: Genetic Markers for Autoimmune Disorder
; FILE REFERENCE: TEX871/4-006US/36000
; CURRENT APPLICATION NUMBER: US/10/067,076
; CURRENT FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: 60/287,893
; PRIOR FILING DATE: 2001-05-01
; PRIOR APPLICATION NUMBER: 09/710,693
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Human Amino Acid
US-10-067-076-1

Query Match
Best Local Similarity 100.0%; Pred. No. 1.1e-21;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 204 EWEQEHRTTCEAFQLOEFDRLTILRNALWVH 234
DB 204 EWEQEHRTTCEAFQLOEFDRLTILRNALWVH 234

RESULT 4
US-10-067-076-20
; Sequence 20, Application US/10067076
; Publication No. US20030104404A1
; GENERAL INFORMATION:
; APPLICANT: Wise, Carol A.
; TITLE OF INVENTION: Genetic Markers for Autoimmune Disorder
; FILE REFERENCE: TEX871/4-006US/36000
; CURRENT APPLICATION NUMBER: US/10/067,076
; CURRENT FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: 60/287,893
; PRIOR FILING DATE: 2001-05-01
; PRIOR APPLICATION NUMBER: 09/710,693
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Human Amino Acid
US-10-067-076-20

Query Match
Best Local Similarity 100.0%; Pred. No. 1.1e-21;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 204 EWEQEHRTTCEAFQLOEFDRLTILRNALWVH 234
DB 204 EWEQEHRTTCEAFQLOEFDRLTILRNALWVH 234
RESULT 5
US-10-067-076-22
; Sequence 22, Application US/10067076
; Publication No. US20030104404A1
; GENERAL INFORMATION:
; APPLICANT: Wise, Carol A.
; TITLE OF INVENTION: Genetic Markers for Autoimmune Disorder
; FILE REFERENCE: TEX871/4-006US/36000
; CURRENT APPLICATION NUMBER: US/10/067,076
; CURRENT FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: 60/287,893
; PRIOR FILING DATE: 2001-05-01
; PRIOR APPLICATION NUMBER: 09/710,693
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 22
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Human Amino Acid
US-10-067-076-22

Query Match
Best Local Similarity 100.0%; Pred. No. 1.1e-20;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 DAFWCRDFTAHRTGYEVLLQRLLDGRKMKCD 38
DB 9 DAFWCRDFTAHRTGYEVLLQRLLDGRKMKCD 38

RESULT 6
US-09-738-626-4016
; Sequence 4016, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO

APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: YATEISHI, NAKO
APPLICANT: SENO, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/577484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO: 4016
LENGTH: 425
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-4016

Query Match 1.9%; Score 8; DB 10; Length 425;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 322 TPSAPAS 329
DB 393 TPSAPAS 400

RESULT 7
US-10-142-231-67
Sequence 67, Application US/10142231
Publication No. US2003007796A1
GENERAL INFORMATION:
APPLICANT: Croteau, Rodney et al.
TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES
FILE REFERENCE: 62773
CURRENT APPLICATION NUMBER: US/10/142,231
CURRENT FILING DATE: 2002-05-08
PRIOR APPLICATION NUMBER: 60/165,250
PRIOR FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 95
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 67
LENGTH: 509
TYPE: PRT
ORGANISM: Taxus cuspidata
US-10-142-231-67

Query Match 1.9%; Score 8; DB 15; Length 509;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 FSGILHGS 318
DB 298 FSGILHGS 305

RESULT 8
US-10-080-170-460
Sequence 460, Application US/10080170
Publication No. US20030129601A1
GENERAL INFORMATION:
APPLICANT: COLE, S.T.
TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
TREATMENT OF MYCOBACTERIOSES
FILE REFERENCE: 03495.0218
CURRENT APPLICATION NUMBER: US/10/080,170

CURRENT FILING DATE: 2002-06-10
PRIOR APPLICATION NUMBER: 60/270,123
PRIOR FILING DATE: 2001-02-22
NUMBER OF SEQ ID NOS: 652
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 460
LENGTH: 1139
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
US-10-080-170-460

Query Match 1.9%; Score 8; DB 16; Length 1139;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 LQRLDLDG 32
DB 191 LQRLDLDG 198

RESULT 9
US-09-879-957-136
Sequence 136, Application US/09879957
Patent No. US20020034755A1
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
HOFFMAN, NO. US20020034755A1b
KAY, Brian K.
FOWLER, Dana M.
MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
USING SAME
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/879,957
FILING DATE: 13-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/630,915
FILING DATE: 03-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 136:
SEQUENCE CHARACTERISTICS:
LENGTH: 57 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: unknown
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 136:
US-09-879-957-136

Query Match 1.7%; Score 7; DB 9; Length 57;
Best Local Similarity 100.0%; Pred. No. 42;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 367 YDYTAON 373
|111111|
Db 9 YDYTAON 15

RESULT 10

US-09-879-957-112
; Sequence 112, Application US/09879957
; Patent No. US20020034755A1
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; HOFFMAN, No. US20020034755A1b
; KAY, Brian K.
; FOMKES, Dana M.
; MCCONNELL, Stephen J.

TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
USING SAME

NUMBER OF SEQUENCES: 227

CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/879,957

FILING DATE: 13-Jun-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/630,915
FILING DATE: 03-Apr-1996

ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie

REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741

TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 112:

SEQUENCE CHARACTERISTICS:
LENGTH: 60 amino acids
TYPE: amino acid

STRANDEDNESS: <unknown>
TOPOLOGY: unknown

MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 112:

US-09-879-957-112

Query Match 1.7%; Score 7; DB 9; Length 60;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 AONSDDEL 377
|111111|
Db 13 AONSDDEL 19

RESULT 11

US-09-864-761-42794
; Sequence 42794, Application US/098644761
; Patent No. US20020046763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
FILE REFERENCE: AeoMica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

SEQ ID NO 42794
LENGTH: 70
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC005412.5
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.97
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.91
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.96
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.92
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3
OTHER INFORMATION: SWISSPROT HIT: P38175, EVALU 4.30e+00

US-09-864-761-42794

Query Match 1.7%; Score 7; DB 9; Length 70;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 ELRSLEE 107
|111111|
Db 26 ELRSLEE 32

RESULT 12

US-09-764-868-1002
; Sequence 1002, Application US/09764868
; Patent No. US2002016871A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT32
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1002
; LENGTH: 111
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-868-1002

Query Match
Best Local Similarity 1.7%; Score 7; DB 10; Length 111;
Matches 7; Conservative 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 LQRLDG 32
Db 53 LQRLDG 59

RESULT 13
US-09-879-957-6
; Sequence 6, Application US/09879957
; Patent No. US20020034755A1
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; HOFFMAN, No. US20020034755A1H
; KAY, Brian K.
; FOWLES, Dana M.
; MCCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; USING SAME
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/879,957
; FILING DATE: 13-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/630,915
; FILING DATE: 03-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: MISTOCK, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-174
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>

US-09-879-957-6
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-879-957-6

Query Match
Best Local Similarity 1.7%; Score 7; DB 9; Length 144;
Matches 7; Conservative 100.0%; Pred. No. 96;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 AONSDQL 377
Db 86 AONSDQL 92

RESULT 14
US-09-860-670-87
; Sequence 87, Application US/09860670
; Patent No. US20020165137A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA127P1
; CURRENT APPLICATION NUMBER: US/09/860,670
; CURRENT FILING DATE: 2001-05-21
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 87
; LENGTH: 155
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (6)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (55)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (101)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (116)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-860-670-87

Query Match
Best Local Similarity 1.7%; Score 7; DB 10; Length 155;
Matches 7; Conservative 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 289 DREVTPL 295
Db 64 DREVTPL 70

RESULT 15
US-09-764-853-550
; Sequence 350, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT206
; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 550
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:

```

: NAME/KEY: SITE
: LOCATION: (76)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (152)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (158)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (165)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (168)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (169)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (173)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: US-09-764-853-550

```

```

Query Match 1.7%; Score 7; DB 9; Length 186;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 365 ALDYTA 371
Db 97 ALDYTA 103

```

Search completed: July 28, 2003, 09:38:52
 Job time : 53 secs

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

[illegible]

```
FT CDS 682..1929
FT /tag= a
FT /product= "PSTPIP"
FT /note= "tyrosine phosphorylated cleavage furrow-
FT associated protein"
XX PN W09835037-1.
XX PD 13-AUG-1998.
XX PF 30-JAN-1998; 98W0-US01774.
XX PR 26-SEP-1997; 97US-0938829.
XX PR 07-FEB-1997; 97US-0798419.
XX PA (GETH ) GENENTECH INC.
XX PI Dowbenko DJ, Lasky LA;
XX DR MPI: 1998-447234/38.
XX DR P-PSDB; AAN71595.
XX PS New PEST-type protein tyrosine phosphatase interacting polypeptide -
XX PT nucleic acids and vectors; for inducing the polymerisation of actin
XX PT monomers in eukaryotic cells and identifying antagonists
XX PS Example 9; Page 62-63; 11pp: English.
XX CC The present sequence encodes murine tyrosine phosphorylated cleavage
XX CC furrow-associated protein (PSTPIP), which is a PEST-type protein
XX CC tyrosine phosphatase (PTP)-interacting polypeptide. PSTPIP induces the
XX CC polymerisation of actin monomers in a eukaryotic cell, by introducing
XX CC a vector containing the nucleic acid sequence encoding PSTPIP into the
XX CC cell. Assays for identifying (ant)agonists of PSTPIP comprise contacting
XX CC PSTPIP with the agent and monitoring the ability of PSTPIP to induce
XX CC actin polymerisation.
SQ Sequence 2100 BP; 548 A; 540 C; 594 G; 418 T; 0 other;

Alignment Scores:
Pred. No.: 3,46e-186 Length: 2100
Score: 2152.00 Matches: 415
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 19 Gaps: 0

US-09-068-377C-1 (1-415) x AAV57973 (1-2100)
QY 1 MetMetAlaGlnLeuGlnPheArGAspAlaPheTrpCysArGAspPheThrAlaHisThr 20
DB 682 ATGATGGCCAGCGTGGATCGAGATGCTTCTGTGGAGGAGCTTACGCGCCACACA 741
QY 21 GlyTyrGlnValLeuLeuGlnArGLeuLeuAspGlyArGlyMetCysLysAspValGlu 40
DB 742 GGGTATGAGAGTCTACTCTGACAGAGGCTGTGGAGCGACAGAAAGATGTGGAG 801
QY 41 GlnLeuLeuArGAlaArGAlaGlnAlaGlnGluArGTrpGlyLysGlnLeuValGlnIle 60
DB 802 GAGCTGCTCAGACAGAGGCGCCAGCGGAGAGAGGTACGGGAAGAGCTGGTGCAGATT 861
QY 61 AlaArGlyAlaGlyLysGlnTrpGlnLeuLeuSerLeuArGTrpPheAspSerLeu 80
DB 862 GCACGCAAGGCTGTGGCCACAGACAGAGATGAATTCCTGTGAGACCTTTGACTCCG 921
QY 81 LysGlnGlnTrpGlnAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaLeuArGlu 100
DB 922 AAGCAGCAAAACAGAGATGTGGCGAGTGCACACATCAGCTGCGCTGGCGCTGAG 981
QY 101 GlnLeuArSerLeuGlnGlnPheArGTrpGlnGlnGlnGlnGlnGlnGlnGlnGln 120
DB 982 GAGCTGGAGGCGCTGGAGAGTTCGACAGAGACAGAAAGAGCGGAAAGATATGAG 1041
```

```
QY 121 AlaIleMetAspArGValGlnLysSerLysLeuSerLeuTrpLysTrpMetLys 140
DB 1042 GCCATATGACCGCTGTCCAGAGAGCAAGTGTCTCTACAGAGACCATGAGATCC 1101
QY 141 LysLysAlaTyrAspGlnLysCysArGAspAlaAspAlaGlnAlaPheGluArG 160
DB 1102 AAGACGCATATGACCAAGAAAGTGCAGGATGCAGATGATGCTGACAGCGCTTCGAGCG 1161
QY 161 ValSerAlaAsnGlyHisGlnLysGlnValGlnLysSerGlnAsnLysAlaLysGlnCys 180
DB 1162 GTGAGTGCCAATGGCCACAGAGCAAGTAGAAAMAAGCCAGAAAGCAAGCAAGCTGC 1221
QY 181 LysGlnSerAlaPheTrpGlnAlaGlnArGValTyrArGAlaAsnIleGlnGlnLeuGluArG 200
DB 1222 AAGGAGTCAAGCCACAGAGGAGAGAGTGTACAGGCAAAATATCCAACTGAGAGAGA 1281
QY 201 AlaArGTrpGlnTrpGlnGlnGlnLysArGTrpTrpCysGlnAlaPheGlnLeuGlnGlu 220
DB 1282 GCGAGACCGAGTGGGAGGAGGAGACCGGACTACTGTGAGGCTTCAGTTCAGAGAG 1341
QY 221 PheAspArGLeuThrIleLeuArGAsnAlaLeuTrpValHisCysAsnGlnLeuSerMet 240
DB 1342 TTTGACCGGCTCACATCTCCCAATGCCCTGTGGGTGACATGTAACAGCTCTCATG 1401
QY 241 GlnCysValLysAspAspGlnLeuTrpGlnGlnValArGLeuThrLeuGlnGlnCysAsp 260
DB 1402 CAGTGTCTCAGAGATATGACCTCTTATGAGAGAGTGCGCTGACCTTGAGGGCTGTGAT 1461
QY 261 ValGlnGlyAspIleAsnGlyPheIleGlnSerLysSerThrGlyArGlnPropAla 280
DB 1462 GTGGAAGGTGACATCATATGCTTCATCCAGTCCAAAGAGCACTGGCAGAGAGCCCAAGCT 1521
QY 281 ProValProCysLysAsnTrpTrpArGAspArGValTrpProLeuIleGlySerProSer 300
DB 1522 CCGGTGCTTATAGACTATGATATGACAGGAGAGTGACCCCATGATGGACCCCTAGC 1581
QY 301 IleGlnProSerCysGlyValIleLysArGPheserGlyLeuLeuHisGlySerProLys 320
DB 1582 ATCCAGCCCTCGTGGCGGTGATAAAGAGGTCTCTGGGCTGTACATGAGATGCCAAG 1641
QY 321 ThrTrpProSerAlaProAlaIleSerThrGlnTrpLeuThrProThrProGluArGAsn 340
DB 1642 ACCACACCTTCTCTCTGCTGTCCACAGAACTGTGCTCCACCCCTGAGCGGAAT 1701
QY 341 GlnLeuValTyrAlaSerIleGlnValGlnAlaTrpGlnGlyAsnLeuAsnSerSerAla 360
DB 1702 GAGTTGCTTACCATTCATCGAAGTGCAGGCGCACCCAGGAAACTTAATCATCAGCC 1761
QY 361 GlnAspTyrArGAlaLeuTyrAspTyrThrAlaGlnAsnSerAspGlnLeuAspIleSer 380
DB 1762 CAGGACTACCGGCGCACTCTACGACTACACTGCACAGAAATTTGATGAGCTGGACATTTCC 1821
QY 381 AlaGlyAspIleLeuAlaValIleLeuGlnGlyGlnAspGlyTrpTrpTrpValGluArG 400
DB 1822 GCGGAGACACTCTGCGGCGTCACTCGGAAGGGAGAGATGCTGTGTGAGATGTGGAGCG 1881
QY 401 AsnGlyGlnArGlyPheValProGlySerTyrLeuGlnLysLeu 415
DB 1882 AAGCAGCAACGTGCTTTGTCCCTGGGTCTACTTGGAGAGAGCTC 1926

RESULT 2
AAA64240
ID AAA64240 standard; DNA; 2100 BP.
XX
XX AAA64240:
AC 20-DEC-2000 (first entry)
XX
XX DNA encoding a murine PTP phosphatase interacting protein (PSTPIP).
DE protein tyrosine phosphatase; PTP; phosphatase interacting protein;
KW PSTPIP; PEST family; protein tyrosine phosphatase; actin monomer;
KW tissue typing; tumour cell; tumour imaging; ss.
```

XX	Mus sp.		
OS			
XX	Key	Location/Qualifiers	
FH	CDS	682..1929	
FT		/tag= a	
FT		/product= "protein tyrosine phosphatase (PNP)	
FT		phosphatase interacting protein (PSTPIP)"	
ET			
XX	US6111073-A.		
PN			
XX	29-AUG-2000.		
PD			
XX	06-FEB-1998;	98US-0020222.	
PF			
XX	07-FEB-1997;	97US-0104590.	
PR			
XX	(GETH) GENENTECH INC.		
PA			
XX	Lasky LA;		
PI			
XX	WPT; 2000-586378/55.		
DX	P-PADB; AAB08486.		
DR			
XX			
XX			
PS	Disclosure; Column 47-50; 48pp; English.		
XX			
CC	The present sequence encodes a protein tyrosine phosphatase (PNP)		
CC	phosphatase interacting protein (PSPPIP). PSPPIP polypeptides are		
CC	bound by and dephosphorylated by the PEST family of protein tyrosine		
CC	phosphataes. PSPPIP associates with actin. PSPPIP is useful for		
CC	inducing the polymerisation of actin monomer in eukaryotic cells by		
CC	introducing the polypeptide into the cell. The polypeptide is useful		
CC	for identifying and isolating PSPPIP homologues in another mammalian		
CC	species, in screening assays to identify antagonists and agonists of		
CC	native PSPPIP polypeptide and as molecular weight markers on protein		
CC	gels. The PSPPIP nucleic acid is useful for tissue typing of specific		
CC	mammalian tissues, for preparing PSPPIP polypeptides by recombinant		
CC	techniques, as hybridisation probes for searching cDNA and genomic		
CC	libraries for the coding sequence of other PSPPIP analogues and to		
CC	isolate homologous genes specifically expressed in tumour cells.		
CC	Antagonists of PSPPIP peptide are useful for inhibiting biological		
CC	activity of the peptide. Antibodies of PSPPIP are useful to identify		
CC	rapidly dividing cells and are used to image tumours comprising such		
CC	rapidly dividing cells.		
XX			
SQ	Sequence 2100 BP: 548 A; 540 C; 594 G; 418 T; 0 other;		
	Alignment Scores:		
	Pred. No.: 3,46e-186 Length: 2100		
	Percent: 2152.00 Matches: 415		
	Best Local Similarity: 100.00% Conservative: 0		
	Query Match: 100.00% Mismatches: 0		
	DB: 21 Gaps: 0		
US-09-068-377C-1 (1-415) x AAA64240 (1-2100)			
OY	1 MetMetAlaGlnLeuGlnIlePheArgAspAlaPheTrpCysArgAspPheThrAlaHisThr 20		
Dd	682 ATGAATGGCCACAGCTGCAGTCCGAAATGCCCTTGCGTCAGGAGCACTTCACGGCCACACA 7411		
OY	21 GlyTYRGIuValLeuLeuGlnArgLeuLeuAspGLYArgLYSMeCYSLYSAspValGlu 40		
Dd	742 GGGTGTGAGGTGCTACTCGACAGAGCGCTGCGACGCGACAGAAGATGTCGAAGATGTGCGAG 8011		
OY	41 GluLeuLeuARGGlnArgGlnAlaGlnIleGluArgTYRGlyGlySGluLeuValGlnIle 60		
Dd	802 GAGCTGCTCAGCACAGAGGGCCAGCGCGGAGGAGAGGTACGCGAAGAGAGCTGTGCAGATT 8611		

QY	61	AlaIrrgYsAlaIagIyGInThrGluMktsnSerLeuArgrSerPheAsnSerLeu	80
Db	862	GCAGCAAGGCTGGGCGCAGACAGAGATGATTCCTCGAGGACCTCTTGACTCCCTG	921
QY	81	LysGInGInThrGluAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaArgGlu	100
Db	922	AAGCAGCAAAACAGAGAAATGTGGCGCATGTCCACATCCAGCTGGCCCTGCGCTCGAG	981
QY	101	GluLeuArgrSerLeuGluGluPheArgrGluArgrGlnLysGlnGlnArgrLysTyrglu	120
Db	982	GAGCTGGCGAGCCTCGAGAGAGTCTCCGAGAGACAGAAAGCAGCCGGAAGAACTATGAG	1041
QY	121	AlaIleuMetAspArgValGlnLysSerLysLeuSerLeuTyrLysLysThMetGlnSer	140
Db	1042	GCCATCATAGGACCGGTGTCCAGAAAGCAGATGTTCGCTCTACAGAAAGAACCATGAGATTC	1101
QY	141	LysLysAlaTyrAspGlnLysCysArgAspAlaAspAspAlaGlnGlnAlaPheGluArg	160
Db	1102	AAGAAAGCATATGACACCAAGACTGCAGAGATCATGATGTGACAGAGCCTTGACACT	1161
QY	161	ValSerAlaAsnGlnLysGlnLysGlnValGluLysSerGlnAsnLysAlaLysGlnCys	180
Db	1162	GTGATGTGCATATGGCCACACAGAACAGATGAAAAAGGCCAGAACAAAGCCAAACACTGTC	1221
QY	181	LysGlnSerAlaThrGlnAlaGlnAlaArgValLysArgGlnAsnIleGlnGlnLeuGluArg	200
Db	1222	AAGGAGTCAGCACAGAGAGCAGAAAGATGTACAGGCAAAATATCGAAACAATCTGAGAGA	1281
QY	201	AlaArgrThrGlnTyrGlnGlnGlnHisArgrThrCysGlnAlaPheGlnLeuGlnGlu	220
Db	1282	CGAGAGACGAGTGGAGACAGAGACCGGAGCTACGTGTGAGGGCTTCCATGTCAGAGAG	1341
QY	221	PheAspArgLeuThrIleLeuArGAsnAlaLeuTyrPValHisCysAsnGlnLeuSerMet	240
Db	1342	TTTGACCGCGCTACACCATCTCCCGGAAATGCCCTGTGGGGCACTGTAAACACTCTTCATG	1401
QY	241	GlnCysValLysAspAspGluLeuTyrGlnGlnValArgLeuThrLeuGlnGlyCysAsp	260
Db	1402	CAGTGTGTCAAGCATGATGACCTCTATGAGAAATGCGGCTGACCTTGAGGGCTGTGAT	1461
QY	261	ValGlnGlyAspIleAsnGlyPheIleGlnSerLysSerThrGlyArgGluProProAla	280
Db	1462	GTGCAAGGTGACATCATCATGAGCTTCATCCAGTCCAAAGACACTGTGCAGAGACCCCGAGCT	1521
QY	281	ProValProTyrGlnAsnTyrTyrAspArgGluValThrProLeuIleGlySerProSer	300
Db	1522	CCGGGCTTTCACGAACCTACTATGACAGGGAGGTGACCCCACTGATTGGCAGCCCTACG	1581
QY	301	IleGlnProSerCysGlyValIleLysArPheSerGlyLeuLeuHisGlySerProLys	320
Db	1582	ATCCAGCCCTCTCTCGGTGTATTAAGAGGTTCTCTGGGCTGCTACATGCAAGTCCAGAG	1641
QY	321	ThrThrProSerAlaProAlaAlaSerThrGlnThrLeuThrProThrProGluArGAsn	340
Db	1642	ACCACACCTTCTGCTCCTGCTGTCTCCACAGAGACTGACTCCACCCCTGAGCGGAT	1701
QY	341	GluLeuValTyrAlaSerIleGluValGlnAlaThrGlnGlyAsnLeuAsnSerSerAla	360
Db	1702	GAGTTGGTCTACGATCCATGAAATGACAGCCAGCCAGGGAACCTTAATCATCTACGCC	1761
QY	361	GlnAspTyrArgAlaLeuTyrAspTyrThrIleGlnAsnSerAspGluLeuAspIleSer	380
Db	1762	CAGAGACTCCGGGCACTCTACGACTACACCTCACAGAAATTTGATGACTGACATTTCC	1821
QY	381	AlaGlyAspIleLeuAlaValIleLeuGlnGlyGluAspGlyTyrPThrValGluArg	400
Db	1822	GCGGGAGACATCTCGCGCTCATCTCGTGAAGGAGAGATGCTGTGTGACTGTGAGACGG	1881
QY	401	AsnGlyGlnArgrGlyPheValProGlySerTyrLeuGlnLysLeu	415
Db	1882	AACGGACACATGGCTTTGTCCCTGGGGCTGACTTGGAGAGAGCTC	1926

AAA07276
ID AAA07276 standard; DNA; 2100 BP.
XX
AC AAA07276;
XX
DT 27-JUN-2000 (first entry)
XX
DE Murine PST phosphatase interacting protein coding sequence.
XX
XX PST phosphatase interacting protein; PSTPIP; tumour therapy;
KM protein tyrosine phosphatase; mouse; ss.
XX
OS Mus sp.
XX
PN US6040437-A.
XX
PD 21-MAR-2000.
XX
PF 29-SEP-1997; 97US-0938830.
XX
PR 07-FEB-1997; 97US-0104590.
XX
PA (GETH) GENENTECH INC.
XX
PI Dowbenko DJ, Lasky LA;
XX
DR WPI: 2000-282393/24.
DR P-PSDB; AAT81928.
XX
PT Novel genes encoding protein tyrosine phosphatase binding proteins
PT useful for isolating homologous genes, e.g. in tumour cells, which
PT provide more specific targets for tumour therapy
PS
PS Claim 5; Column 63-66; 65pp; English.
XX
XX This sequence represents the PST phosphatase interacting protein (PSTPIP)
CC coding sequence of the invention. The protein is a protein tyrosine
CC phosphatase that possesses a non-catalytic domain comprising a proline,
CC serine and threonine rich region and a C-terminal segment of 20 amino
CC acid (aa's) rich in proline, and defines an SH3 binding domain. Nucleic
CC acids encoding native PSTPIP molecules can be used to isolate homologous
CC genes specifically expressed in tumour cells, which might provide more
CC specific targets for tumour therapy. The DNA is also useful for the
CC preparation of PSTPIP polypeptides by recombinant techniques and as
CC hybridisation probes for searching cDNA and genomic libraries for the
CC coding sequence of other PSTPIP polypeptide analogues in other species.
XX
XX Sequence 2100 BP; 548 A; 540 C; 594 G; 418 T; 0 other;
SO

Alignment Scores:
Pred. No.: 3,46e-186 Length: 2100
Score: 2152.00 Matches: 415
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 21 Indels: 0
Gaps: 0
DB:
US-09-068-377c-1 (1-415) x AAA07276 (1-2100)
QY 1 MetMetAlaGlnLeuGlnPheArgAspAlaPheTrpCysArgAspPheThrAlaHisThr 20
DB 682 ATGATGCGCCAGCTGCGATTCGAGATGCTTGTGTCAGGAGCTTCACGGCCACACA 741
QY 21 G1YTYrG1uValLeuLeuGlnArqLeuLeuAspG1ArqLysMetCysLysAspValGlu 40
DB 742 GGGTAAGAGGTGCTACTGTCAGAGCTGCTGAGCGACGAGATGTGCAAGATGTGGAG 801
QY 41 GluLeuLeuArqGlnArqAlaGlnAlaGluGluArqTyrlLysGluLeuValGlnIle 60
DB 802 GAGCTCTCAGACAGAGGGCCGCGGAGAGAGGTACGGGAAGAGCTGTGTCAGATT 861
QY 61 AlaArqLysAlaG1yLglnThrGluMetAsnSerLeuArqThrSerPheAspSerLeu 80
|||||

DB 862 GCACGCAAGCTGCTGCGCCAGACAGATGAATTCCTTGAGGACTCTTGTACTCCCTG 921
QY 81 LysGlnGlnThrGluAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaLeuArqGlu 100
DB 922 AAGCAGCAAAACAGAGATGCGGAGTGCACATCCATGCTGCGCTGCGCTGAG 981
QY 101 GluLeuArqSerLeuGlnGlnPheArqGluArqGlnLysGlnLysGlnLysLysTyrlGlu 120
DB 982 GAGCTCGGAGCCTGAGAGAGTTCGAGAGAGACAGAAAGAGAGGAGGAGAAAGTATGAG 1041
QY 121 AlaIleMetAspArqValGlnLysSerLysLeuSerLeuTyrlLysLysThrMetGluSer 140
DB 1042 GCCATCATGAGCCGTCTCCAGAGAGCAAGTGTCTCTACAAAGAACCATGAGAGTCC 1101
QY 141 LysLysAlaIleTyrlAspGlnLysCysArgAspAlaAspAlaGluGlnAlaPheGluArq 160
DB 1102 AAGAGGCAATATGACAGAGATGTCAGAGATGCAGATGATGTCAGAGGCTTCGAGCGT 1161
QY 161 ValSerAlaAsnGlnLysGlnLysGlnValGluLysSerGlnAsnLysAlaLysGlnCys 180
DB 1162 GTGAGTGCATATGCGCACAGAGCAAGTAAGAAAGAGCCAGAAAGCCAGCAAGTGC 1221
QY 181 LysGluSerAlaThrGluAlaGluArqValTyrlArqGlnAsnIleGlnLysGlnLys 200
DB 1222 AAGAGTCAAGCCACAGAGGCGAAGAGTGTACAGGCAAAATATCGAAACACTGGAGAGA 1281
QY 201 AlaArqThrGluTrpGlnGlnGlnLysIleArqThrTyrlCysGlnAlaPheGlnLysGln 220
DB 1282 GCGAGAGCCGATGGAG 1341
QY 221 PheAspArgLeuThrIleLeuArqAsnAlaLeuTrpValHisCysAsnGlnLeuSerMet 240
DB 1342 TTGACCGGGCTACCATCCCTCCGCAATGCCCTGTGGGTGACATGTAACAGCTCTCATG 1401
QY 241 GlnCysValLysAspAspGluLeuTyrlGlnLysValArgLeuThrLeuGlnLysCysAsp 260
DB 1402 CAGTGTGTCAAGAGATATGAGCTCTATGAGAGAGTGGCTGAGCCCTGTGGGCTGTGAT 1461
QY 261 ValGlnG1yAspIleAsnGlnPheIleGlnSerLysSerThrG1yArqGluPropAla 280
DB 1462 GTGGAAGTGACATCATATGCTTCATCCAGTCCCAAGAGAGAGAGAGAGAGAGAGAG 1521
QY 281 ProValProTrpGlnAsnTyrlTyrlAspArqGluValThrProLeuIleGlnSerProSer 300
DB 1522 CCGGTGCTTATCAGAACTACTATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1581
QY 301 IleGlnProSerCysGlnValIleLysArgPheSerGlyLeuLeuHisGlnSerProLys 320
DB 1582 ATCCAGCCCTCTCGGCTGTGATTAAGAGAGTCTCTGCGCTGTACTATGAGATCCCAAG 1641
QY 321 ThrThrProSerAlaProAlaAlaSerThrGluThrLeuThrProThrProGluArqAsn 340
DB 1642 ACCACACCTTCTCTGCTGCTGCTCCACAGAACTGTACTCCACCCCTGAGCGGAAT 1701
QY 341 GluLeuValTyrlAsnIleGlnValGlnAlaThrGlnGlnLysAsnSerSerAla 360
DB 1702 GAGTGTGCTTACCATTCATCGAAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1761
QY 361 GlnAspTyrlArqAlaLeuTyrlAspTyrlThrAlaGlnAsnSerAspGluLeuAspIleSer 380
DB 1762 CAGAGTACCGGGCAGCTCTGACACTACACTACAGAGAGAGAGAGAGAGAGAGAGAGAG 1821
QY 381 AlaGlyAspIleLeuAlaValIleLeuGlnGluGlnAspGlyTrpTrpThrValGluArq 400
DB 1822 GCGGAGAGACATCTCTGGGCTCATCTCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1881
QY 401 AsnGlyGlnArqGlyPheValIleProGlySerTyrlLeuGlnLysLeu 415
DB 1882 AAGGAGCAACGTGCTTGTCTCTGCTGCTTACTGTGAGAGAGCTC 1926
RESULT 4
AAK89759
ID AAK89759 standard; DNA; 1858 BP.

XX	AAK89759:	
AC		
XX	28-SEP-1999	(first entry)
DT		
XX		
DE	Nucleotide sequence of CD2BP1L.	
XX		
KW	nucleotide; DNA; CDS; CD2; Immune response; Lymphocytes;	
KW	variant; cell adhesion; ds.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	440..1687
FT		/*tag= a
FT		/product= "CD2BP1L"
FT		/note= "Variant of CD2BP1"
XX		
PN	MO9936534-A1.	
XX		
PD	22-JUL-1999.	
XX		
PE	14-DEC-1998;	98MO-US26659.
XX		
PR	13-JAN-1998;	98US-0006428.
XX		
PA	(DAND) DANA FARBER CANCER INST INC.	
XX		
PI	An W, Lh J, Nishizawa K, Reinherz EL;	
XX		
DR	WPI: 1999-444396/37.	
XX		
PT	A human CD2 cytoplasmic tail binding protein, CD2BP1 and related	
PT	polynucleotides, useful in identifying agents which modulate signal	
PT	transduction, cell adhesion or motility	
XX		
PS	Claim 9; Page 98-100; 107pp; English.	
XX		
CC	This is the nucleotide sequence for a variant of CD2BP1, a cdc15-like	
CC	adaptor protein. Anti-CD2BP1 antibodies can be used to assay for the	
CC	presence of CD2BP1 in a cell or tissue sample.	
CC	Modulatory agents identified in the methods are used to inhibit or	
CC	enhance CD2BP1 activity. These agents can be used to modulate	
CC	signal transduction, cell adhesion or motility. CD2BP1 inhibitors are	
CC	useful for enhancing signal transduction, cell adhesion or	
CC	motility. CD2BP1 enhancers have the opposite effect. Mimics of CD2BP1	
CC	protein activity can also be used to modulate signal transduction, etc	
CC	(all claimed).	
CC	Inhibition of CD2BP1 is useful for cancer therapy, to augment the	
CC	immune response to cancer cells and tumors. Enhancement of CD2BP1 is	
CC	useful in immunocompromised individuals and in immunodeficiency related	
CC	diseases. Such diseases and conditions, include rheumatoid arthritis,	
CC	juvenile diabetes, systemic lupus erythematosus, and transplantation, to	
CC	reduce adhesion of the recipient's T cells with tissues in the	
CC	autograft.	
XX		
SO	Sequence 1858 BP; 398 A; 561 C; 594 G; 305 T; 0 other;	
	Alignment Scores:	
	Pred. No.:	1,53e-164
	Score:	1913.50
	Percent Similarity:	92.79%
	Best Local Similarity:	87.98%
	Query Match:	88.92%
	DB:	20
		Gaps: 1
US-09-068-377C-1 (1-415) x AAK89759 (1-1858)		
Oy	1 MetMetLaGlnLeuGlnPheArgAspAlaPheTrpCysArgAspPheThrAlaHisThr 20	
Db	440 ATGATGCCCGACGCTGCAGCTTCACAAAGATGCCCTTTGGTGCAGGAGACTTCACACGCCACACG 495	
Oy	21 G TtYtGtVal Leu LeuGlnArG Leu LeuAspG YArG YMetCysLysAspValG u 40	

QY 340 AsnGluLeuValIYrAlaSerIleGluValGlnAlaThrGlnGlyAsnLeuAsnSer 359
D 1451 AATGAGGGTGTCTACACAGCATCGACAGATACAGGAAACCCGGCTCACCA 1510
QY 360 AlaGlnAspIYrArGAlaLeuIYrAspIYrThrAlaGlnAsnSerAspGluLeuAspIle 379
D 1511 GCCAGAGAGTACCAGGCGCTCTACGATTATACACGAGAACCCAGATGACTGACCTG 1570
QY 380 SerAlaGlyAspIleLeuAlaValIleGluGlnGlyAspGlyIYrIYrThrValGlu 399
D 1571 TCCCGGGAGACATCCTCGAGGATATCTGGAAGGAGAGATGGCTGTGACCTGTGAG 1630
QY 400 ArgAsnGlyGlnArgGlyPheValProGlySerTyrLeuGlnIYrLeu 415
D 1631 AGGAACGGCAGCGTGGCTCTGCTGCTCTACCTGAGAGAGCTT 1678
RESULT 6
AAx89760
ID AAx89760 standard; DNA; 1803 BP.
AAx89760;
AC
XX
DT 28-SEP-1999 (first entry)
DE Nucleotide sequence of CD2BP1s.
XX
KM nucleotide; DNA; CDS; CD2; immune response; lymphocytes;
KM variant; cell adhesion; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 440..1630
FT /tag=a
FT /product="CD2BP1s"
FT /note="variant of CD2BP1"
XX
PN WO936534-A1.
XX
PD 22-JUL-1999.
XX
PF 14-DEC-1998; 98WO-US26699.
XX
PR 13-JAN-1998; 98US-0006428.
XX
PA (DAND) DANA FARBER CANCER INST INC.
XX
PI An W, Li J, Nishizawa K, Reinherz EL;
XX
DR WPI: 1999-444396/37.
XX
PT A human CD2 cytoplasmic tail binding protein, CD2BP1 and related
PT polynucleotides, useful in identifying agents which modulate signal
PT transduction, cell adhesion or motility
XX
PS Claim 11; Page 102-104; 107pp; English.
XX
CC This is the nucleotide sequence for a variant of CD2BP1, a cdg15-like
CC adapter protein. Anti-CD2BP1 antibodies can be used to assay for the
CC presence of CD2BP1 in a cell or tissue sample.
CC Modulatory agents identified in the methods are used to inhibit or
CC enhance CD2BP1 activity. These agents can be used to modulate
CC signal transduction, cell adhesion or motility. CD2BP1 inhibitors are
CC useful for enhancing signal transduction, cell adhesion or
CC motility. CD2BP1 enhancers have the opposite effect. Mimics of CD2BP1
CC protein activity can also be used to modulate signal transduction, etc
CC (all claimed).
CC Inhibition of CD2BP1 is useful for cancer therapy, to augment the
CC immune response to cancer cells and tumors. Enhancement of CD2BP1 is
CC useful in immunocompromised individuals and in immunodeficiency related
CC diseases. Such diseases and conditions, include rheumatoid arthritis,
CC juvenile diabetes, systemic lupus erythematosus, and transplantation, to
CC reduce adhesion of the recipient's T cells with tissues in the

CC autograft.
XX
SQ Sequence 1803 BP; 381 A; 540 C; 585 G; 297 T; 0 other;
Alignment Scores:
Pred. No.: 1.72e-149 Length: 1803
Score: 1748.00 Matches: 341
Percent Similarity: 86.78% Conservative: 20
Best Local Similarity: 81.97% Mismatches: 35
Query Match: 81.23% Indels: 20
D: Gaps: 2
US-09-068-377c-1 (1-415) x AAx89760 (1-1803)
QY 1 MetMetAlaGlnLeuGlnPheArgAspAlaPheIYrPcysArgAspPheThrAlaHisIYr 20
D 440 ATGATGCCCCAGCTCAGTCAATGATGCTTTTGTGTCAGAGACTTACAGCCACACG 499
QY 21 GlyTyrGluValLeuLeuGlnArgLeuLeuAspGlyArgGlyMetCysLysAspValGlu 40
D 500 GGCTACGAGGTGCTCTCAGCGGCTCTGATGTCAGAGAGATGTGCAAGACATGTAG 559
QY 41 GluLeuLeuArgGlnArgAlaGlnAlaGluGluArgTyrGlyLysGluLeuValGlnIle 60
D 560 GAGCTACTGAGGACAGAGCGCCAGCGGAGAGCGGTACGGAGACCTGTGTCAGATC 619
QY 61 AlaArgLysAlaGlyGlyGlnThrGluMetAsnSerLeuArgTyrThrPheAspSerLeu 80
D 620 GCACGGAAGGACAGGTGGCCAGAGGAGATCACTCTGAGGGCTCTTGTACTCTTG 679
QY 81 LysGlnGlnThrGlnLysAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaLeuArgGlu 100
D 680 AAGCAGCAATAGGAAATGTGGCGAGCTCAGCATCCAGCTGACCTGACCTGCTAG 739
QY 101 GluLeuArgSerLeuGlnGluPheArgGluArgGlnLysGlnAlaArgLysIYrGlu 120
D 740 GAGCTGCCGAGTCTCAGAGATTCTGTGAGAGCGCAGAGACAGAGAAAGATATGAG 799
QY 121 AlaIleMetAspArgValGlnLysSerLysLeuSerLeuIYrLysIYrThrMetGluSer 140
D 800 GCCGTATGACACCGGCTCCAGAGACAGAGCTGTCTCTCAAGAGAGCCCTGAGATC 859
QY 141 LysLysAlaTyrAspGlnLysCysArgAspAlaAspAlaGluGlnAlaPheGluArg 160
D 860 AAGAAGACATACAGACAGAGAGTGGCGGAGCGGAGCAGCGGAGACAGCCCTTCAGAGC 919
QY 161 ValSerAlaAsnGlyHisGlnLysGlnValGluLysSerGlnAsnLysAlaLysGlnCys 180
D 920 ATTACGCCCAACGGCCACAGACAGTGTGAGAGAGTCAAGAACCAAGCCAGCATGTC 979
QY 181 LysGluSerAlaThrGluAlaGluArgValIYrArgGlnAsnIleGluGlnLeuGlnIYr 200
D 960 AAGGCTGGCCACCGAGGACAGAGCGGATATACAGAGAGATTGCGAGTGGAGAGAG 1039
QY 201 AlaArgThrGluIYrPGLuGlnGlnHisArgThrCysGluAlaPheGlnLeuGlnIYr 220
D 1040 GTCCGGCTGAGTGGACAGACAGACCGGACACCTGTGAGCCCTTTCAGCTGCAAGAG 1099
QY 221 PheAspArgLeuThrIleLeuArgAsnAlaLeuTyrPvalHisCysAsnGlnLeuSerMet 240
D 1100 TTTGACCCGCTGAGCATTCCTCCGCAACGCCCTGTGGTGTCACAGCAACAGCTTCATG 1159
QY 241 GlnCysValLysAspAspGlnLeuIYrGlnGluValArgLeuThrLeuGlnGlnIYrAsp 260
D 1160 CAGTGTGCAAGAGATGATGATCTACAGAGAGTCCGCTACCTGAGAGCGTGCAGC 1219
QY 261 ValGlnGlyAspIleAsnGlyPheIleGlnSerIYrThrGlyArgGluProProAla 280
D 1220 ATAGAGCGGCAGACATGACAGATTCTATCCAGGCGCAGAGACAGCGGACAGAGCCCGCT 1279
QY 281 ProValProTyrGlnAsnTyrTyrAspArgGluValIYrProLeuIleGlySerProSer 300
D 1280 GAGGTCCGCTTGGCGAGC----- 1297

Db 1237 ----- 1237
QY 302 GlnProSerCysGlyValIleLysArgPheSerGlyLeuHisGlySerProLysThr 321
Db 1238 -----AGGTTCTCTGGACTGCTCAGGAGGATGCCAAGACC 1273
QY 322 ThrProSerAlaProAlaAlaSerThrGluThrLeuThrProThrProGluArgAsnGlu 341
Db 1274 ACTTGCGACGCTTCGTCGGCCACAGAGACCTGACCCCAACCCCGAGGGAATGAG 1333
QY 342 LeuValIleValIleSerTleGluValIleGluAlaThrGluGlyAsnLeuAsnSerAlaGln 361
Db 1334 GGTGCTCAGACAGCATCGACAGTACGAGAGAACCCCGCTCACCACCCAG 1393
QY 362 AspTyrArgAlaLeuTyrAspTyrThrAlaGlnAsnSerAspGluLeuAspIleSerAla 381
Db 1394 GACTACCGGGCCCTTACGATTATACGCGCAACCCAGATGACTGACCTGTCCGG 1453
QY 382 GlyAspIleLeuAlaValIleLeuGluGlyAspGlyTyrPrrThrValGluArgAsn 401
Db 1454 GGAGAC-----ATCCTGGAGGGGAGATGCTGTGACACTGTGAGAGAGAAC 1501
QY 402 GlyGlnArgGlyPheValProGlySerTyrLeuGluLysLeu 415
Db 1502 GGGCAGCGTGGCTTCTCCTGTTCTACTGAGAGAACCTT 1543

RESULT 8

ID AAA52643 standard; DNA; 531 BP.

XX AAA52643;

DT 07-DEC-2000 (first entry)

XX Eosinophil activation cell adhesion peptide gene 1.

KM Eosinophil activation; human; allergy; eosinophilia; cancer;

KM Inflammation; signalling peptide; cell adhesion peptide;

KM G-protein coupled receptor; secreted protein; cell surface protein; ds.

OS Homo sapiens.

PN WO200032630-A2.

PD 08-JUN-2000.

PF 06-DEC-1999; 99WO-US28773.

PR 04-DEC-1998; 98US-0111006.

PA (SEAR) SEARLE & CO G D.

PI Dotson SB, Ma X;

DR WPI: 2000-465041/40.

XX Novel nucleic acids derived from activated eosinophil cells useful for

XX treating allergic diseases such as asthma comprises a specific

XX nucleotide sequence -

XX Claim 1: Page 113-114; 125pp; English.

XX The present invention relates to a number of nucleotide sequences

XX which encode proteins involved in the activation of eosinophils

XX (AA52601-A52679). Eosinophils are involved in immune reactions, and

XX these genes and their proteins provide possible targets for new drugs to

XX combat diseases such as asthma, allergic rhinitis, atopic dermatitis,

XX anaphylaxis, allergic bronchopulmonary aspergillosis, eczema, psoriasis,

XX emphysema, leukemia, lymphomas, ovarian cancer, pneumonia, immune

XX disorders, some connective tissue disorders, and inflammatory conditions

XX including septic shock, arthritis, nephritis, inflammatory bowel disease

XX and Crohn's disease.

XX SQ Sequence 531 BP; 119 A; 157 C; 166 G; 89 T; 0 other;
Alignment Scores:
Pred. No.: 6,54e-60 Length: 531
Score: 756.50 Matches: 147
Percent Similarity: 88.14% Conservative: 9
Best Local Similarity: 83.05% Mismatches: 19
Query Match: 35,15% Indels: 2
Db: 21 Gaps: 1

US-09-068-377C-1 (1-415) x AAA52643 (1-531)

QY 241 GlnCysValIleAspAspGlyLeuTyrGluGluValArgLeuThrLeuGluGlyCysAsp 260
Db 1 CACTGTCTCAGAGATATAGCTTACGAGAGATCGGCTGACGCTGAGAGCTTCAGC 60
QY 261 ValGluGlyAspIleAsnGlyPheIleGlnSerTyrSerThrGlyArgGluProAla 280
Db 61 ATAGACGCCGACATCGACAGTTTCATCCAGGCCACAGAGACGCGCACAGACCCCGCT 120
QY 281 ProValProTyrGlnAsnTyrTyrAspArgGluValThrProLeuIleGlySerProSer 300
Db 121 CCGGTGCCCTTACCAGAACTTACGATCGGAGGTCCACCCCGCTGACACAGACCCCTGCC 180
QY 301 IleGlnProSerCysGlyValIleLysArgPheSerGlyLeuLeuHisGlySerProLys 320
Db 181 ATACAGCCCTCTCGCGCATGATAAAGAGTTCTGTGACTGTGACGAGAGTCCCAAG 240
QY 321 ThrThrPro---SerAlaProAlaAlaSerThrGluThrLeuThrProThrProGluArg 339
Db 241 ACCACTTCTGTGGACACTTCTGTGCTGCTCAGAGAACCTGACCCCAACCCCGAGCGG 300
QY 340 AsnGluLeuValIleValIleSerTleGluValIleGluAlaThrGluGlyAsnLeuAsnSer 359
Db 301 AATGAGGTGCTTACACACCATCCGACGTCAGAGATACAGGAAACCCCGGCTCACA 360
QY 360 AlaGlnAspTyrArgAlaLeuTyrAspTyrThrAlaGlnAsnSerAspGluLeuAspIle 379
Db 361 GCCGAGAGATACCGGGCGCTTCACGATTATACGCCCAACCCAGATAGCTGACCTG 420
QY 380 SerAlaGlyAspIleLeuAlaValIleLeuGluGlyAspGlyTyrPrrThrValGlu 399
Db 421 TCCGCGGAGACATCCTGGAGGTGATCTGGAAGGGGAGAGATGCGTGGAGACTGTGGAG 480
QY 400 ArgAsnGlyGlnArgGlyPheValPro-GlySerTyrLeuGluLysLeu 415
Db 481 AGGAACGGGACGCGTTCCTGCTCCCTGCTTCTACTGAGAAACCTT 529

RESULT 9

ID AAF32684 standard; cDNA; 1580 BP.

XX AAF32684;

DT 22-MAR-2001 (first entry)

XX Human cDNA encoding intracellular signalling molecule INTR447.

XX Human; intracellular signalling molecule; INTRA; immunosuppressive;

XX cytosolic; neuroprotective; nootropic; antiarteriosclerotic; cancer;

XX antiviral; antiparasitic; antihelmintic; antibacterial; antifungal;

XX cell proliferative disorder; arteriosclerosis autoimmune; epilepsy;

XX inflammatory disorder; Addison's disease; gastrointestinal disorder;

XX neurological disorder; Parkinson's disease; Creutzfeldt-Jakob disease;

XX mental disorder; schizophrenia; anxiety; ss.

OS Homo sapiens.

PN WO200077040-A2.

XX 21-DEC-2000.

XX 16-JUN-2000; 2000MO-US16636.
 XX PR 16-JUN-1999; 9905-0139566.
 XX PR 17-AUG-1999; 9905-0149640.
 XX PR 09-NOV-1999; 9905-0164417.
 XX (INCYTE GENOMICS INC.
 XX
 XX Yue H, Tang YT, Hillman JL, Lal P, Bandman O, Baughn MR;
 XX Azimzai Y, Yang J, Reddy R, Lu DAM;
 XX WPI: 2001-025334/03.
 XX P-PSDB: AAB64415.
 XX
 XX New human intracellular signaling molecules, useful for the diagnosis,
 XX prevention and treatment of cell proliferative, autoimmune,
 XX inflammatory, neurological, gastrointestinal, reproductive and
 XX developmental disorders -
 XX
 XX Claim 5; Page 189-190; 192pp; English.
 XX
 XX Sequences AAF32638 - AAF32689 represent cDNA encoding human
 XX intracellular signaling molecules INTRA1 - INTRA52, represented in
 XX AAB64369 - AAB64420. Modulators of the intracellular signaling molecules
 XX of the invention exhibit immunosuppressive; cytostatic; neuroprotective;
 XX neurotropic; antiarteriosclerotic; antiinflammatory; anti-HIV;
 XX neuroleptic; antibacterial; antifungal; antiviral; antiparasitic;
 XX agonists and antagonists are useful for the treatment of a condition
 XX associated with decreased or increased expression of functional INTRA.
 XX Disorders associated with abnormal INTRA expression or activity include
 XX cell proliferative disorders e.g. arteriosclerosis and cancers;
 XX autoimmune or inflammatory disorders e.g. Addison's disease and acquired
 XX immunodeficiency syndrome (AIDS); viral, bacterial, fungal, parasitic,
 XX protozoal and helminthic infections; gastrointestinal disorders e.g.
 XX dysphagia and irritable bowel syndrome; neurological disorders e.g.
 XX epilepsy and Parkinson's disease; prion diseases e.g. Creutzfeldt-Jakob
 XX disease and mental disorders e.g. anxiety, schizophrenia and Tourette's
 XX disorder. Antibodies immuno specific for the INTRA proteins may also be
 XX useful in the diagnosis of the above disorders.
 XX
 XX SQ Sequence 1580 BP; 498 A; 352 C; 372 G; 358 T; 0 other;
 XX
 XX Alignment Scores:
 XX Pred. No.: 2,02e-57 Length: 1580
 XX Score: 736.00 Matches: 144
 XX Percent Similarity: 63.808 Conservative: 64
 XX Best Local Similarity: 44.17% Mismatches: 112
 XX Query Match: 34.20% Indels: 6
 XX DB: 22 Gaps: 3
 XX
 XX US-09-068-377C-1 (1-415) x AAF32684 (1-1580)
 XX
 XX QY 2 MetataglnLeuglnPheargaspalaphetpCyargaspheThralahisthrgly 21
 XX DB 62 ATGAGCGCGTCACTGTTCAAGGAAACTTTGGAGTCAGACATCTCCAGCACATCGGC 121
 XX QY 22 TyrgluValLeuLeuglnArgleuLeuaspGlyArglysmecCyslyssaspValGluGlu 41
 XX DB 122 TTATGACACATTATCCACATCTGAAATGCGCCGCAACACAAAGAGTTTGAAGAC 181
 XX QY 42 LeuLeuArgGlnArgraglnaglnuglnuArgtyGlyLysgluLeuValGlnileala 61
 XX DB 182 TTTCTTAAAGAAAGGAGCAGCAATGGAAGAGGTATGCGAAGACATGCTGCAACCTCTCT 241
 XX QY 62 Arglysalagly---GlylnThrgluMetasSerleuArgthrsrPheaspSerleu 80
 XX DB 242 AGGAAGAACCCGCTGTGACAGTGTGAATCAACACCCCTGAAGGGCCCTTGAAGCTTTC 301
 XX QY 81 LysglnglnThrgluasnValglySerAlahisllleglnleuAlaleuAlaleuArglu 100
 XX DB 302 AAGCAGCAGTAGACAAATGTGGCAATGTCAATTCACCTTGCCAGAGTTTAAAGAA 361

QY 101 GluLeuArgSerleuglnuglnuphearglnuArggluArggluLysgluArglystYrglu 120
 DB 362 GAGCCGAGAGACATGAGACAAATTCAGGAAAGCAAAATCTACACGAAAAAGACAGAG 421
 QY 121 AlailemetaspArgValGlnLysSerLysLeuSerleuThryLysLysThrmelGuser 140
 DB 422 CTCATATGATGATGCTCATTAACCAAGAGCTTCAATTCACAGAAACCATGATGCA 481
 QY 141 LysLysAlaTyraSpolnLysCysArgaspAlaaspAlaGlnGlnAlaPheGluArg 160
 DB 482 AAGAACACTATGAGCAGGAATGCCGAGCAAGAGATGAGCAGAAAGCCCTGACCGG 541
 QY 161 ValSerAlaAsnGlylscInLysGlnValGluLysSerGlnAsnLysAlaLysGlnCys 180
 DB 542 AGTCCCAACCTGTGTGACACCCGAGACACAGAAAGAACTTTGTGAACTGCGCACTTCA 601
 QY 181 LysGluSerAlaThrglnuArgValTyraArgGlnAsnllleglnLeuGlnuArg 200
 DB 602 AAGACCGCAGTAGAGGACTCAGACAAAGCATCATCTCGACATCGCACCCCTGATAG 661
 QY 201 AlaArgThrgluTrpGlnGlnGlnuArgthThrcysGlnuAlaPheGlnLeuGlnu 220
 DB 662 GTCCGAGAGAGTGGCAGAGTACATCAAGCCCTGGAGCATTTAGGCTCAAGAA 721
 QY 221 PheaspArgLeuThrlleuArgasnAlaLeuTrpValHisCysasnGlnLeuSermet 240
 DB 722 TGTGACGATATAACTCTTCCGATGATGATGTGTGTACATGTGATGATGATGATGAT 781
 QY 241 GlnCysValLysaspGlnuLeuTyrgluGlnuValArgLeuThrgluGlnuGlnu 260
 DB 782 CAATGTGTACACAGTGTGATTAATGTGCAAGATGCGAAGAGTGTAAAGATGTGCAC 841
 QY 261 ValGlnGlyAspLysasnGlyPheleGlnSerLysSerThrclyArgGlnuPropoAla 280
 DB 842 ATTCAAGAGGACTTAAATACCTTTGTAATCAACGCAAACTGACAGATTCACACAGCA 901
 QY 281 ProvalProtyrGlnAsnTyraSpArgGlnuValThrProleuIlegLysSerProser 300
 DB 902 CCCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 961
 QY 301 IleGlnProSerCysGlyValIleLysArgPheSerGlyLeuLeuHisGlySerProlys 320
 DB 962 ACAGGGCCTAAC-----TTGCCAAGAGAGAGACCCCTC-----CCAATTCCTAAA 1006
 QY 321 ThrThrProSerAlaPro 326
 DB 1007 AGCTCACAGATGATGCC 1024
 RESULT 10
 ID AAI59033 standard; cdna; 1606 BP.
 AC AAI59033;
 XX
 XX 22-OCT-2001 (first entry)
 XX
 XX Human polynucleotide SEQ ID NO 1236.
 DE
 XX Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 XX peripheral nervous system; neuropathy; central nervous system; CNS;
 XX Alzheimer's; Parkinson's disease; Huntington's disease; hemostatic;
 XX amyotrophic lateral sclerosis; Shy-Drager syndrome; chemotactic;
 XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 XX leukemia; ss.
 XX
 XX Homo sapiens.
 OS
 XX
 XX MO200153312-AL.
 PN
 XX
 XX 26-JUL-2001.
 PD
 XX
 XX 26-DEC-2000; 2000MO-US34263.
 PF

XX 21-JAN-2000: 2000US-0488725.
 PR 25-APR-2000: 2000US-0552317.
 PR 09-JUL-2000: 2000US-0598042.
 PR 19-JUL-2000: 2000US-0620312.
 PR 03-AUG-2000: 2000US-0653450.
 PR 14-SEP-2000: 2000US-0662191.
 PR 19-OCT-2000: 2000US-0693036.
 PR 29-NOV-2000: 2000US-0727344.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Dmanec RT;

DR WPI: 2001-442253/47.
 DR P-PSDB: AAM39877.

PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 PS Claim 1; SEQ ID NO 1236; 10078pp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AA42213) with nootropic.
 CC immunosuppressant and cyostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localized neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/Inhibin activity, chemoclastic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

XX Sequence 1606 BP; 519 A; 350 C; 375 G; 362 T; 0 other;

XX Alignment Scores:

Pred. No.: 2,06e-57 Length: 1606
 Score: 736.00 Matches: 144
 Percent Similarity: 63.808 Conservative: 64
 Best Local Similarity: 44.17% Mismatches: 112
 Query Match: 34.20% Indels: 6
 DB: Caps: 3

US-09-068-377C-1 (1-415) x AA159033 (1-1606)

QY 2 MetAlaGlnLeuGlnIlePheArgAspAlaPheTrpCysArgAspPheThrAlaHisThrGly 21
 DB 58 ATGACCGCGCTCACTGTTCAAGGAAACTTTTGGAGTGCACAGACATCTCCACACCATCGGC 117
 QY 22 TrrGluValLeuLeuGlnAlaGlnLeuAspGlyArgGlyMetCysLysAspValGluGlu 41
 DB 118 TATGACAAACATTTCCAAACATCTGAACAAATGGCCGCAAGCAACATGTTTGAAGAC 177
 QY 42 LeuLeuArgGlnArgAlaGlnAlaGlnGluArgGlyGlyLysGluLeuValGlnIleAla 61
 DB 178 TTTCTTAAAGAAAGCCGACCAATTTGAAGAGAGATGTGGCAAAAGCTGCTCAACCTCTCT 237
 QY ArgLysAlaGly--GlyGlnThrGluMetAsnSerLeuArgThrSerPheAspSerLeu 80
 DB 238 AGAAGAAGCCGTGTGACAGCTCTGAATAATCAACACCTTGAAGCGCCCTTGAAGCTTTC 297
 QY 81 LysGlnGlnThrGlnAsnValGlySerAlaHisIleGlnIleLeuAlaLeuAlaArgGlu 100
 DB 298 AAGCAGCAAGTGAACAATGTGGCAACAATGTCAATTCAGCTTGCAAGAGTTTAAAGAGA 357

QY 101 GlnLeuArgSerLeuGlnGluPheArgGlnArgGlnLysGlnArgLysLysTyrglu 120
 DB 358 GAGGCCAGGAAAGATGGAAGAAATTTAGGGAAAAAGCAAAACTACACCAAAAAAGACAGAG 417
 QY 121 AlaIleMetAspArgValGlnLysSerLysLeuSerLeuTyrlLysTrhMetGluSer 140
 DB 418 CTCATATATGATGCTATCCATTAACCAAAAGAGCTTACCAATTCAGAAACCATGATGACA 477
 QY 141 LysLysAlaTyrlAspGlnLysCysArgAspAlaAspAlaGlnGlnAlaPheGluArg 160
 DB 478 AAGAAGACATATGAGCAAAATGCGGAGCAAAATAGGACGAACAGCCGCTCAGCCG 537
 QY 161 ValSerAlaAsnGlnLysGlnLysGlnValGlnLysSerGlnAsnLysAlaLysGlnCys 180
 DB 538 AGTCCCAACTGGTGAACCCGAGACACACAGAAAGCTTTTGTGAACCTGGCACTTCA 597
 QY 181 LysGlnSerAlaThrGlnAlaGlnArgValTyrlArgGlnAsnIleGlnGlnGluArg 200
 DB 598 AAGACCCAGTAGAGACTCAGACAAAGCATACATGCTGCACATCGGCACCCCTGATTAAG 657
 QY 201 AlaArgThrGluTrpGluGlnGlnHisArgThrThrCysGlnAlaPheGlnGlnGlu 220
 DB 658 GTCCGAGAGAGAGTGGACAGTGCACATCAAGGCTCGAGGCAATTGAGGCTCAGAA 717
 QY 221 PheAspArgLeuThrIleLeuArgAsnAlaLeuTrpValHisCysAsnGlnLeuSerMet 240
 DB 718 TGTCAGAGATTAACCTTCTCCGSAATGCATTTGCTTACTTGTCAATCAGCTGCACAA 777
 QY 241 GlnCysValLysAspAspGluLeuTyrlGluGlnValArgLeuThrLeuGlnGlyCysAsp 260
 DB 778 CAATGTCTCACCAGTATGAAATGTAGCAACATCCGAAAGAGTTTGAAGATGTGCAGC 837
 QY 261 ValGlnLysPheIleAsnGlnPheIleGlnSerLysSerThrGlyArgGluProProAla 280
 DB 838 ATTAGAGGCGCATTAATTAATCTTGTGAATCAACGCAAAATGTGACAGATTCCACGACA 897
 QY 281 ProValProTyrlGlnAsnTyrlTyrlAspArgGluValThrProLeuIleGlySerProSer 300
 DB 898 CCCATCATGATATGAAATTTCTACTCTCCCGAAGAAATGCAAGTCCAGCAGAAAGGCT 957
 QY 301 IleGlnProSerCysGlyValIleLysArgPheSerGlyLeuLeuHisGlySerProLys 320
 DB 958 ACAGGCGCTAAC-----TTGCAAGAGAGAGACCCCTC-----CCAAATTCCTAAA 1002
 QY 321 ThrThrProSerAlaPro 326
 DB 1003 AGCTCACCAGATGATCC 1020
 RESULT 11
 ID AA160819 standard; cDNA; 1630 BP.
 AC AA160819;
 XX 22-OCT-2001 (first entry)
 DT 22-OCT-2001 (first entry)
 XX Human polynucleotide SEQ ID NO 4808.
 DE Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amotrophic lateral sclerosis; Shy-Drager Syndrome; chemoclastic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 XX Homo sapiens.
 OS Homo sapiens.
 XX MO200153312-A1.
 PN MO200153312-A1.
 XX 26-JUL-2001.
 PD 26-JUL-2001.
 XX 26-DEC-2000; 2000WO-US34263.
 PF 26-DEC-2000; 2000WO-US34263.
 XX

PR 21-JAN-2000: 2000US-048725.
PR 25-APR-2000: 2000US-0552317.
PR 09-JUL-2000: 2000US-0598042.
PR 19-JUL-2000: 2000US-0620312.
PR 03-AUG-2000: 2000US-0653450.
PR 14-SEP-2000: 2000US-0662191.
PR 19-OCT-2000: 2000US-0693036.
PR 29-NOV-2000: 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RR;
XX
XX WPI: 2001-442253/47.
DR P-PSDB; AAM41663.
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Claim 1; SEQ ID NO 4808; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX Sequence 1630 BP; 511 A; 363 C; 386 G; 370 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 2,1e-57 Length: 1630
Score: 736.00 Matches: 144
Percent Similarity: 63.80% Conservative: 64
Best Local Similarity: 44.17% Mismatches: 112
Query Match: 34.20% Indels: 6
DB: 22 Gaps: 3
US-09-068-377C-1.(1-415) x AA160819 (1-1630)
QY 2 MetAlaGlnLeuInpHeaRgAspAlaPhetRcysArGAspHeThraAlaHisThrcly 21
DB 94 ATAGCGGCTCACTGTTCAGAGAACTTTGGAGTGCAACATCTCAGACCCATCGGC 153
QY 22 TyGluValLeuLeuGlnArGLeuLeuAspGlyArGlyMetCysLysAspValGluGlu 41
DB 154 TATGACACACTTATCCACATCTGAGACAAATGGCCGCAAGACCTCAAGATTGAGAGAC 213
QY 42 LeuLeuArGlnArGAlaGlnAlaGluGluArGTYrGlyLysGluLeuValGlnIleAla 61
DB 214 TTCTTAAAGAAAGGCGCAATTCGAAGAGAGATGCGAAAGATCTGCTCAACCTCTCT 273
QY 62 ArGlyAlaGly---GlyGlnThrGluMetAsnSerLeuArGTrHrSerPheAspSerLeu 80
DB 274 AGGAAGAAAGCCGTGTGGACAGTCTGAATACACACCTGGAAGCGGCTTGAAGTCTTC 333
QY 81 LysGlnGlnThrGluAsnValGlySerAlaGlyIleGlnLeuAlaLeuAlaLeuArgGlu 100
DB 334 AAGCAGCAAGTATACAAATGTGGCACAATGTCACTTACGCTTGACAGAGATTGAGAGAA 393
QY 101 GluLeuArGSerLeuGluGluInpHeaRgGluArGAlaGlnLysGluGlnArGlyLysTYrGlu 120

DB 394 GAGGCCAGCAAGATGGAAGATTCAGGAGAAAGCAAAACTACAAAGCAAAAGACAGAG 453
QY 121 AlaIleMetAspArGValGlnLysSerLysLeuSerLeuTYrLysLysTrMetGluSer 140
DB 454 CTCATATGATGATGATTCATTAACAAAGACCTTACAAATTCAGAAACCATGATGATGA 513
QY 141 LysLysAlaTYrAspGlnLysCysArGAspAlaAspAlaGlnGluGlnAlaPheGluArg 160
DB 514 AAGAGACACTATGAGCAAAATGCCGGACAAAGATGAGCAGACAGCGCCGTACGCGG 573
QY 161 ValSerAlaAsnGlnHisGlnLysGlnValGluLysSerGlnAsnLysAlaLysGlnCys 180
DB 574 AGTGCCAACTGTGGAGACCCGAAAGCAAGAAAGCTTTTGTGAAACCTGGCAACTTCA 633
QY 181 LysGluSerAlaThrGluAlaGluArgValTYrArGlnAsnIleGluGlnLeuGluArg 200
DB 634 AAGACCGCAGTACAGACTGACAAAGCATPACTGCTGACATCGCCACCTGGATTAAG 693
QY 201 AlaArGThrGluTrpGluGlnGluHisArGThrThrcysGluAlaPheGlnLeuGlnGlu 220
DB 694 GTCCGAGAAAGTGGCAGATGAGACATCAAGSCCTCGCAGGCATTGAGGCTCAGAA 753
QY 221 PheAspArGLeuThrIleLeuArGAsnAlaLeuTrpValHisCysAsnGlnLeuSerMet 240
DB 754 TGTGAAAGCAATAACTTCTCCGGAATGCAATTGGTTACATGTGAATCAGCTGTCAAA 813
QY 241 GlnCysValLysAspAspGluLeuTYrGluGluValArGLeuTrpLeuGlnGluCysAsp 260
DB 814 CAATGTCTCCACAGTGAAGATGTACGAACAGCTCCGAAGACATTGAGAAATGTGGAGC 873
QY 261 ValGluLysAspIleAsnGlyPheIleGlnSerLysSerThrcysGluArgGluProAla 280
DB 874 ATTCAAGGGACATGTAATCTTGTGAATCAACGCAAACTGACAGCATTCACACAGCA 933
QY 281 ProValProTYrGlnAsnTYrTYrAspArGluValThrProLeuIleGlySerProSer 300
DB 934 CCCATCATGTATGAGATTTCTACTCTCCAGAAAGATCAGTCCACGACGAGAAAGCT 993
QY 301 IleGlnProSerCysGlyValIleLysArGpHeSerLysLeuLeuHisGlySerProLys 320
DB 994 ACAGGCGCTAAC-----TTGGCAAGAGAGAGACCCTC-----CCAATTCTTAA 1038
QY 321 ThrThrProSerAlaPro 326
DB 1039 AGCTCACCAGATGATGCC 1056
RESULT 12
AAS27373 standard; cDNA; 914 BP.
ID AAS27373
XX
XX AAS27373:
AC 07-NOV-2001 (first entry)
DE
XX cDNA encoding novel signal transduction pathway protein, Seq ID 408.
KW Neuroprotective; cytoskeletal; dermatological; immunosuppressive; tumour;
KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;
KW immune system disorder; rheumatoid arthritis; inflammatory condition;
KW organ transplant rejection; infection; hepatitis C; blood disorder;
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
KW reproductive system; gastrointestinal; liver disorder; AIDS; ss;
KW acquired immune deficiency syndrome.
XX
XX Homo sapiens.
OS
XX
XX W0200154733-A1.
PN
XX
XX 02-AUG-2001.
PD

XX 17-JAN-2001; 2001WO-US01312.
 PF XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0203515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 14-JUL-2000; 2000US-0217496.
 PR 26-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 14-AUG-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225477.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226688.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234277.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 01-NOV-2000; 2000US-0241826.
 PR 08-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249246.
 PR 17-NOV-2000; 2000US-0249247.
 PR 17-NOV-2000; 2000US-0249257.
 PR 17-NOV-2000; 2000US-0249259.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Barash SC, Ruben SM;
 PI WPI; 2001-465460/50.
 XX P-PSDB; AAU17456.
 DR

XX Novel polypeptides useful for diagnosing, treating, preventing and/or
 PT prognosing disorders related to the proteins, including cancers, immune
 PT disorders and neuronal disorders
 XX
 PS Claim 1: SEQ ID NO 408; 880bp; English.

XX The invention relates to novel isolated polypeptides (I), and
 CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
 CC diagnosing, preventing and treating diseases including immune system
 CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
 CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
 CC transplant rejections and graft versus host disease, infectious diseases
 CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
 CC other blood-related disorders (sickle cell anaemia), myeloproliferative
 CC disorders, primary haematopoietic disorders, hyperproliferative
 CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
 CC disorders (e.g. Alzheimer's disease and Parkinson's disease), chromosomal
 CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
 CC disorders (e.g. glomerulonephritis), cardiovascular disorders
 CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
 CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
 CC Addison's disease), reproductive system disorders, gastrointestinal
 CC disorder (inflammatory disorders), liver disorders (cirrhosis),
 CC as stimulators of B-cell responsiveness to pathogens, activators of
 CC T-cells, to induce higher affinity antibodies, and as a means to induce
 CC tumour proliferation in pathologies e.g. acquired immune deficiency
 CC syndrome (AIDS). AAS26976-AAS27850 represent novel signal transduction
 CC pathway protein coding sequences and PCR primers of the invention.
 XX

Alignment Scores:
 Pred. No.: 4,1e-39 Length: 914
 Score: 531.50 Matches: 123
 Percent Similarity: 62.50% Conservative: 52
 Best Local Similarity: 43.93% Mismatches: 100
 Query Match: 24,70% Indels: 8
 DB: 22 Gaps: 3

US-09-068-377C-1 (1-415) x AAS27373 (1-914)

QY 2 MetAlaGlnLeuGlnPheArgAspAlaPheTrpCysArgAspPheThrAlaHisThrGly 21
 DB 80 ATGACGCGCTCACTGTTCAAGGAAACTTTTGGAGTGCAGACATCCTCAGCACCATCGGC 139
 QY 22 TyrGluValLeuLeuGlnArgLeuLeuAspGlyArgIleuMetCysIleuAspValGluGlu 41
 DB 140 TATGACAACTTATCCAACTGGAACATGAGCCGCAAGAACTGCAAGAGTTTGAAGAC 139
 QY 42 LeuLeuArgGlnArgAlaGlnAlaGluArgTyrGlyIleuGluLeuValGlnIleAla 61
 DB 200 TTTCCTAAAGAAAGGCGACGAATTTGAAGAGGTATNGGCAAAATCGCTCAACCTCTCT 259
 QY 62 ArgIleuAlaGly--GlyGlnThrGluMetAsnSerLeuArgThrSerPheAspSerLeu 80
 DB 260 AGGAACAACCCGTGTGACAGTCTGAAATCAACACCCCTGAGCGGCCCTTGAAGTCCTTC 319
 QY 81 LysGlnGlnThrGlnAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaLeuArgGlu 100
 DB 320 AAGCAGCAAGTACACATGTGGCACAATGTCACATTCAGCTTCACAGAGTTTAAGAGAA 379
 QY 101 GluLeuArgSerLeuGlnGluPheArgGluArgGlnLysGlnGlnArgLysLysTyrGlu 120
 DB 380 GAGGCGAGAGATGAGAGATTCAGGAGAAACAAATAATACAGCAAGAAAGAGACGAG 439
 QY 121 AlaIleuMetAspArgValGlnLysSerLysLeuSerLeuTyrLysLysThrMetGluSer 140
 DB 440 CTCMTATGTCGTCTATCCMTTAAACAAAGAGTTTCAATTCAGAAACCAATGAGATGCA 499
 QY 141 LysLysAlaIleTyrAspGlnLysCysArgAspAlaAspAspAlaGlnGlnAlaPheGluArg 160
 DB 500 AAGAGAACTATGAGCAGAAATGCCGGGACAAAGATGAGCAGACAGAGCCGTCACGCCG 559

QY 161 ValSerAlaAsnGlyHisGlnLysGlnValGluLysSerGlnAsnLysAlaLysGlnCys 180
 DB 560 ACTGCCAACCTGCTGTGAACCGAAGCAACAGAAAGCTTTT-TGCAACTGGCAACTTCA 618
 QY 181 LysGlnSerAlaThrGluAlaGluArgValTyrArgGlnAsnIleGlnLeuGluArg 200
 DB 619 AAGACCGCAGTAGAGACTACAGCAAAAGCATTCATCTGCACATCGCACCCTGGATTAAG 678
 QY 201 AlaArgThrGluTrpGluGlnGlnHisArgThrThrCysGluAla-PheGlnLeuGlnI 220
 DB 679 GTCCGGAAGAGTGGCAGAGTACGACATCAAGCCCTGGAGCATTTTGAAGCTTCAAGA 738
 QY 220 UpheAspArgLeu-ThrIleLeuArgAsnAlaLeuTrpValHisCysAsnGlnLeuSer 240
 DB 739 ATGTGACCAATTAATCTTTCCGGAATGCATTTGGTGTACATGTGATCA-CTGTAC 797
 QY 240 eteGlnCysValLysAspAspGluLeuTyrGluGluValArgLeuThrLeuGluGlyCys 260
 DB 798 AACAAATGTGTACCCGT-GATGAATGTACAGACCA--TNCNAGAGTTAGAAATGTGCA 853
 QY 260 spValGluGluLysPheIleAsnGlyPheIleGlnSerLysSerThrGlyArgGluPro 278
 DB 854 GCATTCAANAAGGCGATT---TGATCTTGGAATNAGCAAACTGCGACATTCGCC 906

RESULT 13

ID AAH16224 standard; cDNA; 2591 BP.

AAH16224;

26-JUN-2001 (first entry)

Human cDNA sequence SEQ ID NO:15042.

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

Homo sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUL-2000; 2000EP-0116126.

29-JUL-1999; 99JP-0248036.

27-AUG-1999; 99JP-0300253.

11-JAN-2000; 2000JP-0118776.

02-MAY-2000; 2000JP-0183767.

09-JUN-2000; 2000JP-0241899.

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

Ishii S, Sugiyama T, Wakamatsu A, Negai K, Otsuki T;

WPI: 2001-318749/34.

Claim 8: SEQ ID 15042; 2537bp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a


```

Db      1 AAACCTACACGAAAAAGACAGACCTCATATGATGCTATCCATAAACAAGAGGCTTA 60
Qy      133 LeuYrlyslsYThrMetGlsrLysLysAlaTyrsplnLysCysArgrAspAlaasp 152
Db      61 CAATTCAAGAAACCATGATGATGCAAGAAAGACTATGAGCAAAATGCCGGAGCAAAAGAT 120
Qy      153 AspaIagluGlnAlaPheGluArValSerAlaasnGlnLysGlnValGluVal 172
Db      121 GAGCAGACAGCGCGCTGACAGGAGAGTGGCAACTGTGGAGACCCGAGACCAAGAAAG 180
Qy      173 SerAlaasnLysAlaLysGlnCysLysGluSerAlaThrGlnAlaGluArValTyra 192
Db      181 CTTTGTGAAACCTGCAACTTCAAAAGCCGACTAGAGAGCTGAGCAAAAGCATACATG 240
Qy      193 GlnasnIleGluGlnLysGluArValAlaThrGlnLysGlnValGlnVal 212
Db      241 CTGACACATCGGACCCCTGGATAGGTCGAGAGAGTGGCAAGTGGAGTGGACATCAAGCC 300
Qy      213 CysGlnAlaPheGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 232
Db      301 TGGCAGCATTTGAGGCTCAAGATGTAAGCAATTAACCTTCCGGAATGCATTTGTCG 360
Qy      233 ValHisCysasnGlnLysSerMetGlnCysValLysAspSgIuLysGlnLysGln 252
Db      361 TTACATGTGATGATGCTGTCACAAACATGTCTCCAGCATGTAATGTAAGCAAGATC 420
Qy      253 ArgLeuThrLeuGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 272
Db      421 CGAAGAGATTAGAAATGTGACAGATTCAGAGGACATGTGATGTAATCAAGCC 480
Qy      273 SerThrGlnArgLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 292
Db      481 AAAACTGACAGATTCACACAGCAGCCATCATGATGATTTCTACTCTCCCAAG 540
Qy      293 ThrProLeuIleGlnLysSerProSerIleGlnLysSerGlnLysValLysVal 312
Db      541 AATGCACTCCACAGGAAAGGCTACAGGCGCTAAC-----TTGGCAGAGAGAGA 591
Qy      313 GlyLeuLeuHisGlnLysSerProLysThrThrProSerAlaPro 326
Db      592 CCCTT-----CCAATTCCTAAAGCTNACAGATGATCCC 627

RESULT 15
ABL87423/c
ID      ABL87423 standard; cDNA: 488 BP.
XX
AC      ABL87423:
XX
DT      17-MAY-2002 (first entry)
XX
DE      Human ovarian cancer related cDNA clone SEQ ID NO:10401.
XX
KM      Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX
OS      Homo sapiens.
XX
PN      WO200192581-A2.
XX
PD      06-DEC-2001.
XX
PF      29-MAY-2001: 2001MO-US17756.
XX
PR      26-MAY-2000: 2000US-207484P.
XX
PA      (CORI-) CORIXA CORP.
XX
PI      Algate PA, Harlocker SL, Jones R;
XX
DR      WPI: 2002-122075/16.
XX
PT      Composition for therapy and diagnosis of ovarian cancer comprising
        polypeptide of a ovarian tumor polypeptide, polynucleotide encoding

```

```

PT      polypeptide, antibody specific to polypeptide or T cell expressing
XX      polypeptide
PS      Claim 1: SEQ ID 10401; 489bp; English.
XX
CC      The present invention describes a composition: (I) comprising: carriers
CC      and immunostimulants; and a polypeptide (II) having a ovarian tumor
CC      polypeptide encoded by a polynucleotide (III) having a cDNA sequence
CC      (S1) from the 10912 nucleotide sequences as given in ABL77023 to
CC      ABL87934, (III) encoding (II) having a sequence (S2), a T cell
CC      population of (II), or antigen presenting cells that express (II).
CC      (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to
CC      (S1) can be used for detecting ovarian cancer in a patient's biological
CC      sample preferably serum or ovarian tissue. The method comprises
CC      contacting a biological sample from a patient with (IV), detecting the
CC      amount of polynucleotide hybridising to (IV) and comparing the amount to
CC      a predetermined cutoff value and thereby detecting ovarian cancer in the
CC      patient, where the amount of polynucleotide hybridising to (IV) is
CC      detected preferably by polymerase chain reaction (PCR). (I) comprising
CC      (III) and/or (II) is useful for stimulating and/or expanding T cells
CC      specific for an ovarian tumor protein comprising contacting T cells
CC      with (III) or (II). (III) is useful in design and preparation of
CC      ribozyme molecules for inhibiting expression of the tumor polypeptides
CC      and proteins in tumor cells; and to isolate a full length gene from a
CC      suitable library e.g., a tumor cDNA library using well known
        techniques.
XX
SQ      Sequence 488 BP: 77 A; 147 C; 139 G; 123 T; 2 other:
XX
Alignment Scores:
Pred. No.: 1,53e-25 Length: 488
Score: 378.50 Matches: 81
Percent Similarity: 83.33% Conservative: 4
Best Local Similarity: 79.41% Mismatches: 16
Query Match: 17.59% Indels: 2
DB: 24 Gaps: 1
XX
US-09-068-377c-1 (1-415) x ABL87423 (1-488)
Qy      315 LeuHisGlnLysSerProLysThrThrPro--SerAlaProAlaAlaSerThrGluThrLeu 333
Db      488 CTGCNCGAGGATGCCAAGACCACTTGTGGCAGCTTTGNTGCTGCACAGAGACCTG 429
Qy      334 ThrProThrProGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 353
Db      428 AC-CCACCCCGAGCGGATGAGGTGCTTACACAGCCATGCCATGACGAGATACAG 370
Qy      354 GlyAsnLeuAsnSerSerAlaGlnAspTyraGlnAlaLeuTyraSpTyThrAlaGlnAsn 373
Db      369 GGAACCCCGGCTCCACAGCCAGGAGTACCGGCGCTTACGATTATACAGCGCAGAAC 310
Qy      374 SerAspGlnLeuAsnLysSerAlaGlnAspTyraGlnAlaLeuTyraSpTyThrAlaGlnAsn 393
Db      309 CCAGATGAGCTGACCTGTGCGGGGAGACATCTCTGAGGTGCTGGAAGGGAGGAT 250
Qy      394 GlyTrpTrpThrValGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 413
Db      249 GGCTGTGAGCTGTGAGAGAGAAAGCGGACGCTTGTGCTGCTGCTGCTGCTGCTGCTG 190
Qy      414 LysLeu 415
Db      189 AAGCTT 184

```

Search completed: August 3, 2003, 12:19:31
 Job time : 346 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: August 3, 2003, 12:07:26 ; Search time 1853 Seconds

(without alignments)
3627.160 Million cell updates/sec

Title: US-09-068-377C-1

Perfect score: 2152

Sequence: 1 MMAQLQFRDAFWCRDFTAH.....WTVERNCGRGFPVPSYLEKL 415

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 segs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgnr_1/USPRO_pool/US09068377/runat_28072003.091316.15326/app_query.fasta.1.583
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45
-DOCLALIGN=200 -THR_SCORE=Pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09068377 @CGN 1.1 1506 @runat.28072003.091316.15326 -MCPU=6 -ICPU=3
-NO_MMAP -LARGEROUTRY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estlm:*
5: em_estlov:*
6: em_estlpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estlm:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1336.5	62.1	888	13	B1556340
2	1149	53.4	1022	14	BM922694
3	1089.5	50.6	828	12	BM819706
4	1074	49.9	629	12	BE853802
5	1056	49.1	698	12	BE676651
6	1030.5	47.9	797	12	BG820623
7	1010	46.9	766	13	B1904988
8	1001	46.5	594	14	B0553004
9	982.5	45.7	1120	14	BM924109
10	953.5	44.3	811	13	B1912313
11	889	41.3	543	14	BM933148
12	876	40.7	502	9	AA038079
13	871	40.5	507	9	A1322422
14	855	39.7	733	14	BQ207763
15	833	38.7	737	13	B1910444
16	813.5	37.8	578	13	B1774480
17	804.5	37.4	653	12	BE282676
18	802.5	37.3	700	9	AL581028
19	773	35.9	708	13	B1906059
20	764	35.5	483	10	AM659798
21	719	33.4	1139	13	BM563898
22	711	33.0	455	13	BM087768
23	711	33.0	584	12	BF020734
24	699.5	32.5	447	12	BF549426
25	684	31.8	568	12	BM221823
26	683	31.7	485	13	BE878305
27	682	31.7	495	10	AW916545
28	637.5	29.6	641	9	A1952795
29	636.5	29.6	467	12	BF707133
30	623.5	28.0	454	13	BM107477
31	621	28.9	542	9	A1324866
32	614	28.5	547	10	AM540915
33	565	26.3	381	13	B1029004
34	555	25.8	944	14	BQ720739
35	548	25.5	785	13	B1911163
36	548	25.5	870	14	BQ900279
37	545.5	25.3	355	12	BF554556
38	545.5	25.3	518	14	BQ190642
39	543	25.2	1038	14	BQ051519
40	539	25.0	560	10	AV616620
41	539	25.0	774	14	BQ605464
42	536.5	24.9	559	14	BQ029347
43	531.5	24.7	396	10	AM407004
44	531	24.7	939	12	BF530009
45	524	24.3	1209	13	BM547486

ALIGNMENTS

RESULT 1
LOCUS B1556340
DEFINITION 603237461F1 NCI_CGAP_Mam3 Mus musculus CDNA clone IMAGE:5290479 5',
LOCUS B1556340
ACCESSION B1556340
VERSION B1556340
KEYWORDS mrna sequence.
SOURCE EST.
ORGANISM house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE

JOURNAL
COMMENT

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1734 row: j column: 16
High quality sequence stop: 783.

FEATURES
source

Location/Qualifiers
1. 888
/organism="Mus musculus"
/strain="129,C57BL/6J,FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5290479"
/clone_lib="NCI_CGAP_Mam3"
/tissue_type="tumor, gross tissue"
/dev_stage="10 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: Salt; Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference for transgenic model: Xu et al., Nature Genetics 22, 37-43 (1999)."
BASE COUNT 241 a 227 c 263 g 157 t
ORIGIN

Alignment Scores:

Pred. No.: 3.14e-123 Length: 888
Score: 1336.50 Matches: 273
Percent Similarity: 85.62% Conservative: 1
Best Local Similarity: 85.31% Mismatches: 20
Query Match: 62.11% Indels: 26
DB: 13 Gaps: 3

US-09-068-377C-1 (1-415) x B1556340 (1-888)

QY 85 GIUAASVAGLISeralHisIleGlnleuAlaIleuArgGluGluIleuArgSer 104
DB 2 GAGAAITGIGGCGAGGACATCCAGCGCCCTGCGCTCGAGAGAGCTGGCGAGC 61
QY 105 LeuGluGluPheArgGluArgGlnLysGlnArgLysLysTyrGluAlaIleMetasp 124
DB 62 CTGGAGAGAGTTCGAGAGAGAGAGAGAGAGAGAGAGAGATAGAGCCATCATGAC 121
QY 125 ArgValGlnLysSerLysLeuSerLysLysLysLysLysLysLysLysLys 144
DB 122 CGTGTCCAGAT 181
QY 145 AspGlnLysCysArgAspAlaAspAspAlaGlnGlnAlaPheGlnLysValSerAlaSn 164
DB 182 GACCAAGAGTGCAGGAGATGATGATGCTGAGCGGCTTCGAGCGTGTGAGTCCAT 241
QY 165 GlyHisGlnLysGlnValGlnLysSerGlnAsnLysAlaLysGlnCysLysGlnSerAla 184
DB 242 GGGCCAG 301
QY 185 ThrGluAlaGluArgValTyrArgGlnAsnIleGlnLysGlnLysGlnLysGln 204
DB 302 ACAGAGCGAG 361
QY 205 TrrpGlnGlnLysHisArgThrThrCysGlnAlaPheGlnLysGlnLysGlnLys 224
DB 362 TGGGAGCGAG 421
QY 225 ThrIleLeuArgAsnAlaLeuTrpValHisCysAsnGlnLysSerMetGlnCysValLys 244
DB 422 ACCATCTCCGCAATGCCCTGTGGGTGACTGTAAACAGCTCTCCATGAGTGTCAAG 481

QY 245 AspAspGluLeuTyrGluGluValArgLeuThrLeuGlnGlyCysAspValGluLysp 264
DB 482 GATGATGAGCTCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 541
QY 265 IleAsnGlyPheIleGlnSerLysSerThrGlyArgGluProAlaProValProTyr 284
DB 542 ATCAATGCGCTTCATCCAGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 601
QY 285 GlnAsnTyrTyrAspArgGluValThrProLeuIleGlySerProSerIleGlnProSer 304
DB 602 CAGAACTACTATGACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 661
QY 305 CysGlyValIleLysArgPheSerGlyLeuLeuHisGlySerProLysThrThrProSer 324
DB 662 TCGGAGTGATATAAGAGGTCCTCTGGGCTGCCTACATGGAAGTCCAGACACACCTTCT 721
QY 325 AlaPro-AlaAlaSerThrGluThrLeuThrProThrProGluArgAsnGluLeuValTyr 344
DB 722 GCTCCTTGTCT-----GAGACTCCGACTCCCAACCTGAGCGAGATGAGTGTCTA 772
QY 344 rAlaSerIleGluValGlnAlaThrGlnGlyAsnLeuAsnSer--SerAlaGlnAspTyr 363
DB 773 CCATTCCTCATGCAATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 832
QY 364 ArgAlaLeuTyrAspTyrThrAlaGlnAsnSerAspGluLeuAspIleSerAlaGlyasp 383
DB 833 CGGCGACTCTACGAT---AACTGAGAGAGATCTGATGAG----- 868
QY 384 IleLeuAlaValIleLeuGlnGlyLysAspGlyTyrTrpThrValGluArgAsn 401
DB 869 -----TGCACTTCCGAGAGAT 886

RESULT 2
BM922694 1022 bp mRNA linear EST 12-MAR-2002
LOCUS BM922694 5', mRNA sequence.
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1022)
NIH-MGC http://mgs.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1795 row: f column: 19
High quality sequence stop: 634.

FEATURES
source

Location/Qualifiers
1. 1022
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5755794"
/clone_lib="NIH-MGC_118"
/tissue_type="Leukocyte"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6; Site:1: NotI; Site:2: EcoRV (destroyed); RNA source leukocytes from anonymous pool of non-activated adult donors. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range 1.2-3.3 kb. Library is normalized and enriched for

Full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 027. Note:
this is a NIH-MGC Library."

BASE COUNT 236 a 282 c 353 g 151 t
ORIGIN

Alignment Scores:

	2.11e-104	Length:	1022
Pred. No.:	1149.00	Matches:	247
Percent Similarity:	88.63%	Conservative:	18
Best Local Similarity:	82.61%	Mismatches:	23
Query Match:	53.39%	Indels:	11
	14	Gaps:	0

US-09-068-377C-1 (1-415) x BM922694 (1-1022)

```

OY 1 MetMetAlaGlnLeuGlnPheArGAspAlaPheTtPCysArGAspPheThrAlaHisThr 20
Db 133 ATGATGCCCGACGCTCAGTCAAGATGCTTTGGTGCAGGAGCTTCACAGCCACAGC 192
OY 21 G1YTYrG1uVal1LeuLeuGln1nArG1eUeUAspG1ArG1yMeTcYsAspValG1u 40
Db 193 GGCTACGAGGTGCTCTGCGCGCTTCTGGATGCGAGAGATGTCAAAGACATGAG 252
OY 41 GluLeuLeuArG1nArG1AlaG1nAgluG1uArG1yG1yLysG1uLeuValG1nle 60
Db 253 GAGCTACGAGGACAGAGGCCCGAGCGAGCGGTACGAGAGAGAGAGAGAGAGAGAG 312
OY 61 AlaArG1ySA1aG1yG1nThrCluMetAsnSerLeuArG1ThrSerPheAspSerLeu 80
Db 313 GCACGAGAGGACGAGTGGCGAGAGAGATCAACTCCTGAGGCGCTCTTGACTCTTG 372
OY 81 LysGlnGlnThrGlnAsnValG1ySerAlaHis1leGlnLeuAlaLeuArG1u 100
Db 373 AAGCAGCAAAAGGAAATGTGGCGAGCTCACACATCCAGTCCCTGACCTGCGCTGAG 432
OY 101 GluLeuArG1SerLeuG1u1nArG1yG1uArG1nLysG1uArG1yLys1yTrG1u 120
Db 433 GAGTGGCGGAGTCTCGAGAGAGTTCGTGAGAGAGAGAGAGAGAGAGAGAGAGATAG 492
OY 121 Ala1leMetAspArG1ValG1nLysSerLysLeuSerLeuTyLysLysThrMetG1uSer 140
Db 493 GCCGTCATGAGACGGGCTCCAGAGAGAGAGAGTGTGCTCTACAGAGAGAGAGAGAG 552
OY 141 LysLysAlaTyArSpG1nLysCysArGAspAlaAspAspAlaG1nAglu1nA1nAheG1uArG 160
Db 553 AAGAGACATACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 612
OY 161 ValSerAlaAsnG1yHisG1nLysG1nValG1uLysSerG1nAsnLysAlaLysG1nCyS 180
Db 613 ATTACGCGCAACGGCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 672
OY 181 LysG1uSerAlaThrG1uAgluArG1yValTyArG1nAsn1leG1uG1nLeuG1uArG 200
Db 673 AAGGACTGCGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 732
OY 201 AlaArG1ThrG1uTPG1uG1nLys1nArG1ThrCysG1uValA1nAheG1nLeuG1nle 220
Db 733 GTCCGCGGCTGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 792
OY 220 nPheAspArG1LeuThr1leLeuArGAsnAlaLeuTPVal1nLysCysAsnG1nLeuS 239
Db 793 GTTTCACGCGGCTGACATTTCTCCGCAACCCCTGGGTGGAGAGAGAGAGAGAGAG 852
OY 239 erMeG1nCySValLysAspAspG1uLeuTyTrG1uG1nValArG1eUeThrLeuG1u 258
Db 853 CCATGCACTGGGCTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 912
OY 258 YCysArpValG1uG1yAsp1leAsnG1yPhe1leG1nSerLysSerThrG1yArG1u 277
Db 913 CTGAGAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 972
OY 278 -ProProAlaPro-ValProTyTrG1nAsnTyTrTyr 288

```

Db 973 CCCCCCGTCCGCGGCGCCCTACGAAACTTTT 1007

RESULT 3
Bg819706
LOCUS
DEFINITION
Bg819706 828 bp mRNA linear EST 22-MAY-2001
602783012F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4933682
5', mRNA sequence.
ACCESSION
Bg819706
VERSION
Bg819706.1 GI:14167293
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 (bases 1 to 828)
NIH-MGC <http://mgi.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLM10861 row: h column: 03
High quality sequence stop: 767.
Location/Qualifiers
1..828

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4933682"
/clone_lib="NCI_CGAP_Brn67"
/tissue_type="anaplastic oligodendroglioma with 1p/19q
loss"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: brain; Vector: PCMV-SPORE6; Site_1: NCI;
Site_2: Salt; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 195 a 239 c 263 g 130 t 1 others
ORIGIN

Alignment Scores:

	1.4e-98	Length:	828
Pred. No.:	1089.50	Matches:	222
Percent Similarity:	88.76%	Conservative:	15
Best Local Similarity:	83.15%	Mismatches:	24
Query Match:	50.63%	Indels:	7
	12	Gaps:	1

US-09-068-377C-1 (1-415) x BG819706 (1-828)

```

OY 155 GluGlnAlaPheGluArG1ValSerAlaAsnG1yHisG1nLysG1nValG1uLysSerG1n 174
Db 1 GAGCAGCGCTTCCAGGCGCATTAAGCGCCACAGAGAGAGAGAGAGAGAGAGAGAGAG 60
OY 175 AsnLysAlaLysG1nCyS1yG1uSerAlaThrG1uAgluArG1yValTyArG1nAsn 194
Db 61 AACAAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
OY 195 l1eG1uG1nLeuG1uArG1nArG1ThrG1uTPG1uG1nLys1nArG1ThrCysG1u 214
Db 121 ATTGCCACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
OY 215 AlAheG1nLeuG1nLupheAspArG1eUeThr1leUeArGAsnAlaLeuTPVal1nL 234
Db 181 GCTTTTCAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
OY 235 CysAsnG1nLeuSerMetG1nCySValLysAspAspG1uLeuTyTrG1uG1nA1nAheG 254

```

Db		241	AGCAACCAAGCTCCATGTCAGTGTGCACAAGATGATGACTTACAGAGAAGTGCGGCTG	300
Oy		255	ThrlengluclycysaspvalglugelyaaipllleasnclgPheIleelinserlyserrTr	274
Db		301	ACCGTTGAAGGCTGCCAGCATAGACGCCGACATGCAGATTTCATCCAGGACCAGAACAG	360
Oy		275	GlyArgGluproProalaproValproTYrglnasnrTYrtyAspargluValThrpo	294
Db		361	GGCAGACAGACCCCCCGCTCCGGGCGCCCTACAGAACTATTAGATGGGAGGTCACCCC	420
Oy		295	LeulleGlySerProserIleGlnProSerCyglyVallleYsarPheserGlyeu	314
Db		421	CTGACACGACGACCCCTGGCATACAGCCGCTCCGCGCATATAAGAGGTTCTCTGACCTG	480
Oy		315	LeuhlsGlySerProLysThrThrpo---SeralProAlAlaserthrGluThrleu	333
Db		481	CTGCACAGGAAGTCCAGAACCACTTCGTGGCAGCTTCGTGCTGCCACAGAACCCCTG	540
Oy		334	ThrProThrProGluArGasnGluLeuValTYrAlaSerlleGluValGlnlaThrGln	353
Db		541	ACCCCCANCCCCGAGCGAATGAGGGTGTCTACACAGCATCGCAGTGAGAGATACAG	600
Oy		354	GlyAsnLeuasnsrSerIlaGlnaspTyraTalaaleuTYrAsPTyrThralaInsn	373
Db		601	GGAAAACC-GGCTCACACACCCAGAGATRCGGGGCGCTCTACGATTATACACCCGACAC	659
Oy		374	SeraspgluLeuaSPIseerIlaGlyspIleleuAlaVallleleGluGlygu-As	393
Db		660	CCAGATGAGACTGGACCTGCCCGGGAGACATCCTGAGAGTATCTCTGGAAAGGGAAGA	719
Oy		393	P-GlyTrpThrphrhalGlu-ArGasnGlylnaTarglglyPheValPro-GlyserTYle	412
Db		720	TtGGTTGGTGACCTGTGACGAGGAACGGGACGCTGTGCTTCCTCCGTGGTTCCTACT	779
Oy		412	U-GluLysLeu	415
Db		780	GGGAGAGACTT	790
RESULT 4				
BE853802				
LOCUS				
DEFINITION		BE853802	629 bp mRNA linear EST 27-SEP-2000	
ACCESSION			Ux3Jf1.y1 Soares_thymus_2NBMT Mus musculus cDNA clone	
VERSION			IMAGE:3472365 5' similar to FR:P97814 P97814 PEST PHOSPHATASE	
KEYWORDS			INTERACTING PROTEIN ; mRNA sequence.	
SOURCE			BE853802.1 GI:10312372	
ORGANISM			EST.	
JOURNAL			house mouse.	
COMMENT			Mus musculus	
REFERENCE			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutelestomi;	
AUTHORS			Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
TITLE			NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.	
			National Cancer Institute, Cancer Genome Anatomy Project (CGAP).	
			Tumor Gene Index	
			unpublished (1997)	
			Contact: Robert Strausberg, Ph.D.	
			Email: ccapbs-remail.nih.gov	
			This clone is available royalty-free through LINT ; contact the	
			IMAGE Consortium (info@image.llnl.gov) for further information.	
			MG1:191725	
			Seq primer: -40RP from Gldco	
			High quality sequence stop: 466.	
FEATURES			Location/Qualifiers	
SOURCE			1..629	
			/organism="Mus musculus"	
			/strain="C57BL/6J"	
			/db_xref="taxon:10090"	
			/clone="IMAGE:3472365"	
			/clone_idb="Soares_thymus_2NBMT"	
			/sex="male"	
			/tissue_type="Thymus"	

[illegible]

ACCESSION mRNA sequence.
VERSION BG676651
KEYWORDS BG676651.1 GI:13908048
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 698)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: James Cleaver, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLAM10599 row: O column: 12
High quality sequence start: 3
High quality sequence stop: 661.
Location/Qualifiers
1. 698
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:4748387"
/clone_lib="NCI_CGAP_Skn4"
/tissue_type="squamous cell carcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 175 a 177 c 245 g 100 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 2,46e-95 Length: 698
Score: 1056.00 Matches: 207
Percent Similarity: 94.40% Conservative: 12
Best Local Similarity: 89.22% Mismatches: 13
Query Match: 49.07% Indels: 1
DB: 12 Gaps: 0
US-09-068-377C-1 (1-415) x BG676651 (1-698)
QY 8 ArgAspAlaPheTrpCysArgAspPheThrAlaHisThrGlyTyrGluValLeuLeuGln 27
Db 2 CAAGATGCGCTTTGGTGGCAAGGACTTCACAGCCACAGGCGTACGAGTGTCTCTCAG 61
QY 28 ArgLeuLeuAspGlyArgGlyMetCysLysAspValGluGluLeuLeuArgGlnArgAla 47
Db 62 CGGCTTCTGTGATGGCAGGAAGATGTCTCAAGACATGAGGAGCTACTGAGGAGAGAN-GCC 120
QY 48 GlnAlaGluGluArgTyrGlyLysGluLeuValGlnIleAlaArgLysAlaGlyGln 67
Db 121 CAGGCGAGAGAGGGTACGCGGAAGAGCTGTCTCAGATCCAGGAGGAGGAGGCGCAG 180
QY 68 ThrGlnTAsnSerLeuArgThrSerPheAspSerLeuLysGlnInThrGluAsnVal 87
Db 181 ACGGATCAACACCTCGAGGGGCTCTTGTACTCTTGAAGCAGCAATAGGAAATGTG 240
QY 88 GlySerAlaHisIleGlnLeuAlaLeuAlaLeuArgGluLeuArgSerLeuGluGln 107
Db 241 GCGAGCTCACATCCAGCTGCGCTGACCTCGTGAAGAGCTCGAGACTCTCGAGAG 300
QY 108 PheArgGluArgGlnLysGluGlnArgLysLysTyrGlnAlaIleMetAspArgValGln 127
Db 301 TTTCGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 360

QY 128 LysSerLysLeuSerLeuTyrLysLysThrMetGlnSerLysLysAlaTyrAspGlnLys 147
Db 361 AAGAGCAAGCTGTGCTCTTACAGAGGCGCATGGAGTCCAGAGACTATGACAGAGAG 420
QY 148 CysArgAspAlaAspAspAlaGluGlnAlaPheGluArgValSerAlaAsnGlyHisGln 167
Db 421 TGCCGGAGCGCGGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 480
QY 168 LysGlnValGluLysSerGlnAspLysAlaLysGlnCysLysGlnSerAlaThrGluGln 187
Db 481 AAGCAGGTGGAGAGAGTACAGACAGACCGAGGAGTGCAGAGCTGGCCACCGAGGCA 540
QY 188 GluArgValTyrArgGlnAsnIleGlnLeuGlnLeuGlnArgAlaArgThrGluTrpGluGln 207
Db 541 GAGGGGTATTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 600
QY 208 GlnHisArgThrThrCysGluAlaPheGlnLeuGlnGluPheAspArgLeuThrIleLeu 227
Db 601 GAGCAGCGGAGCCAGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 660
QY 228 ArgAsnAlaLeuTrpValHisCysAsnGlnLeuSer 239
Db 661 CGCAACGCCCTGTGGTTGACAGCAACCAAGCTCTCC 696
RESULT 6
BG820623 797 bp mRNA linear EST 22-MAY-2001
DEFINITION 60278291F1 NCI_CGAP_Brn67 Homo sapiens cdna IMAGE:4933731
LOCUS 5', mRNA sequence.
ACCESSION BG820623
VERSION BG820623.1 GI:14168210
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 797)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: David N. Louis, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLAM10861 row: J column: 04
High quality sequence stop: 787.
Location/Qualifiers
1. 797
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:4933731"
/clone_lib="NCI_CGAP_Brn67"
/tissue_type="anaplastic oligodendroglioma with 1p/19q
loss"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: brain; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 188 a 225 c 260 g 123 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 1,05e-92 Length: 797
Score: 1030.50 Matches: 216
Percent Similarity: 87.83% Conservative: 15
Best Local Similarity: 82.13% Mismatches: 30
Query Match: 47.89% Indels: 6

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Plate: LLM11649 row: j column: 05
High quality sequence start: 6
High quality sequence stop: 764.

FEATURES	Location/Qualifiers
source	1. .766

```

/organism="Mus musculus"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone="IMAGE:5257060"
/clone_1ib="NCI_CGAP_Lu33"
/tissue_type="pooled lung tumors"
/lab_host="PH10B (phage-resistant)"
/notes="Organ: lung; Vector: pRTT30-Pac (Pharmacia) with a
modified polylinker; Site 1: NotI; Site 2: EcoRI; 1st
strand cDNA was prepared from mRNA obtained from pooled
lung tumors with a NotI primer (5'
GGTTACCAATCTGAGAGTGGAGCGCCCTCTCTTTTATTTT 3').
Double-stranded cDNA was ligated to Eco. RI adaptors
(Pharmacia), digested with NotI and cloned into the NotI
I and Eco RI sites of the modified pRTT3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "

```

Alignment Scores:	
Pred. No.:	Length:
Score:	1,11e-90
Percent Similarity:	1010.00
Best Local Similarity:	98.53%
Query Match:	98.04%
DB:	46.93%
	13
US-09-068-377C-1 (1-415) x B1904988 (1-766)	

[illegible]

DB:	14	Gaps:	0
US-09-068-377C-1 (1-415) x BQ553004 (1-594)			
OY	17	ThralaHISThG1TyrGluValLeuLeuGlnAArgLeuAspG1ARgylsMetCys	36
Db	1	ACGGCCACACAGGGATATGAGTGTCTACTGCGAGAGGCTGCTGGACGCGAGAGATTGTGC	60
OY	37	LysAspValGluGluLeuLeuArgGlnAArgAlaGlnAlaGluGluArgTyrG1LysGlu	56
Db	61	AAGATGTGGAGAGACTGCTCGACAGAGAGGGCCAGGGCGAGAGAGATGACGGAGAGAG	120
OY	57	LeuValGln1IleAlaArgLysAlaGlyGlnThrGluMetAsnSerLeuArgThrSer	76
Db	121	CTGGTGCAAGATTGCACGCAAGGCTGGTGGCCAGACAGAGATGAATTCCTCAGGACCTCC	180
OY	77	PheAspSerLeuLysGlnIleThrGluAsnValG1SerAlaHis1IleG1LeuAlaLeu	96
Db	181	TTTGACTCCTCGAAGCAGCAAGAGATGTGGCAGTGCACACATCCAGCTGGCCTG	240
OY	97	AlaLeuArgGluGluLeuArgSerLeuGluGluPheArgGluArgGlnLysGluGlnArg	116
Db	241	GCCCTGCGAGAGACTCGGAGCCTGGAGGAGTTCGAGAGAGACAAAGAGCAGCGG	300
OY	117	LysLysTyrGluAla1IleMetAspArgValGlnLysSerLysLeuSerLeuTyrLysLys	136
Db	301	AAGAAGTATGAGGCCATCATGACCGTGTCCAGAAAGACCAAGTTGTCTCTACAAAG	360
OY	137	ThrMetGluSerLysLysAlaTyrAspGlnLysCysArgAspAlaAspAlaGluGln	156
Db	361	ACCATGCGAGTCCAGAGAGCATATGACCAAGAGTGCAGAGATGCGATGATGCTGACGAG	420
OY	157	AlaPheGluArgValSerAlaAsnGlyHisGlnLysGlnValGluLysSerGlnAsnLys	176
Db	421	GCCTTCGAGCGGTGAGTGGCCAAATGGCCAGCAAGCAAGTAAAGAGCAGACAGCAA	480
OY	177	AlaLysGlnCysLysGluSerAlaThrGluAlaGluArgValTyrArgGlnAsn1IleGlu	196
Db	481	GCCAAAGCAGTCCAAAGAGTCAAGCCCAAGAGGCGAGAAAGTGTACAGCAAAATATCGAA	540
OY	197	GlnLeuGluArgValArgThrGluThrPheGluGlnGluVH1SAATGThrCysGlu	214
Db	541	CAACTGGAGAGAGCCAGGACCGAGTGGAGAGGAGCAGCGAGTACTGTGAG	594
RESULT 9			
BM924109	1120 bp	mRNA	linear
LOCUS	AGENCOURT_6630376	NIH_MGC_116	Homo sapiens cDNA clone IMAGE:5760315
DEFINITION	5', mRNA sequence.		
ACCESSION	BM924109		
VERSION	BM924109.1	GI:19374488	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 1120)		
JOURNAL	NIH-MGC http://mhc.nci.nih.gov/ .		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgabbs-r@mail.nih.gov		
	Tissue Procurement: Life Technologies, Inc.		
	cDNA Library Preparation: Life Technologies, Inc.		
	DNA Sequencing by: Agencourt Bioscience Corporation		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNL at:		
	http://image.lnl.gov		
	Plate: L1AM12807 row: c column: 04		
	High quality sequence stop: 668.		
FEATURES	Location/Qualifiers		
source	1..1120		

/organism="Homo sapiens"
 /db.xref="taxon:9606"
 /clone="IMAGE:5760315"
 /clone_lib="NIH_MGC_116"
 /lab_host="DH10B"
 /note="Organ: pooled colon, kidney, stomach; Vector:
 PCMV-SPO6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
 source anonymous pool of 3 colons, age 26 yo male, 49 yo
 female, 71 yo male colon; 46 yo male kidney, and pool of 2
 stomachs, 62 yo male and 70 yo female. Library is
 oligo-dT primed and directionally cloned (EcoRV site is
 destroyed upon cloning). Average insert size 1.4 kb,
 insert size range 1-3 kb. Library is normalized and
 enriched for full-length clones and was constructed by C.
 Gruber (Invitrogen). Research Genetics tracking code
 023. Note: this is a NIH_MGC Library."

BASE COUNT 248 a 316 c 373 g 183 t

ALIGNMENT SCORES:

Pred. No.: 1,06e-87 Length: 1120
 Score: 982.50 Matches: 211
 Percent Similarity: 73.44% Conservative: 24
 Best Local Similarity: 65.94% Mismatches: 47
 Query Match: 45.66% Indels: 38
 DB: 14 Gaps: 8

US-09-068-377c-1 (1-415) x BM924109 (1-1120)

QY 1 MetMetAlaGlnLeuGlnPheArgSPAlaPheTrpCysArgAspPheThrAlaHisThr 20
 Db 199 ATGATGCCCGCAGCTGCAGATCAAGATGCTTTGGTGACAGGACTTCACAGCCACACG 258
 QY 21 GLyTrpGluValLeuLeuGlnArgLeuLeuAspGlyArgLysMetCysLysAspValGlu 40
 Db 259 GGCCTACAGAGTGCTGCTGACGGCTTCTGATGGCAGGAAGATGCAAGATGGAG 318
 QY 41 GluLeuLeuArgGlnAlaGlnAlaGlnGluArgTrpGlyLysGluLeuValGlnIle 60
 Db 319 GAGCTACTGAGGACGAGGAGGCGCCAGGAGGAGCGGTACGGAGAGAGCTGTGACAGTC 378
 QY 61 AlaArgLysAlaGlyGlnThrGluMetAsnSerLeuArgThrSerPheAspSerLeu 80
 Db 379 GCAGGAGGACGAGTGCGCCAGACGAGATCAATCCCTGAGGGCTCTTGAATCCTTG 438
 QY 81 LysGlnGlnThrGluAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaLeuArgGlu 100
 Db 439 AAGCAGCAAAATGGGAATGTGGGACGCTCACACATCCAGCTGGCCCTGACCTGCTGAG 498
 QY 101 GluLeuArgSerLeuGlnGluPheArgGluArgGlnLysGlnArgLysLysTyrGlu 120
 Db 499 GAGCTGGGAGTCTCGAGAGTCTCGTGAAGGACAGCAAGGAGCAAGAACTATGAG 558
 QY 121 AlaIleMetAspArgValGlnLysSerLysLeuSerLeuTyrLysLysThrMetGluSer 140
 Db 559 GCCCTCTTGGACCGGGTCCAGAAAGCAAGCTGCTCTCAAGAGGCGCATGAGATGCC 618
 QY 141 LysLysAlaTyrAspGlnLysCysArgAspAlaAspAlaGlnGlnAlaPheGluArg 160
 Db 619 AAGAGACATACGACGAGAGTGGGAGCGGAGCAGCGGAGAGGCGCTTGAGAGCGC 678
 QY 161 ValSerLysAsnGlnLysGlnValGlnLysSerGlnAsnLysAlaLys-GlnCys 180
 Db 679 ATTAGCCCAAGCGCCACCAAGACAGTGGAGAAAGTCCGACAAAGACGAGGAGCTG 738
 QY 180 sLysGluSerAlaThrGluAlaGluArgValTyrArgGlnAsnIleGlnGlnLeuArg 200
 Db 739 CAAGGACTCGGCCACCGACGACGAGCGGTATACAGCAGAGCATTCGCGCTGAGAA 798
 QY 200 gAlaArgThrGluTrpGlnGlnLysArg---ThrThrCysGlnAlaPheGlnLeuG 219
 Db 799 GGTCCGGGCTGAGTGGAGCAGGAGCACCAGGACCAACTGGTGAAGGCTTTTCACTGCA 858

QY 219 ngluphe---AspArgLeuThrIleLeuArgAsnAlaLeuTrpValHisCysAsnGlnLe 238
 Db 859 AGAATTTTGACCGGGCTGACACTTCTTCCCAAC----- 892
 QY 238 uSerMetLeuGlnCysValLysAspAspGluLeuTyrGlnGluValArgLeuThrLeuGlu 258
 Db 893 -----CGCTCCGGTGGTTCACACAGCAGCAGCAGTCCCGCCAGCAAG 936
 QY 258 yCysAspValGlnGlnLysPheLeuAsnGlyPheIleGlnSerLysSerThrGlyArgGly 277
 Db 937 G-----GGGTCAAGATGATGAGCTTCTACCC-----AGGGAAT 972
 QY 278 -ProPro-----AlaProValProTyrGlnAsnTyr 287
 Db 973 TCCCGCTTGACCCCTTGGAAGCGCTGCACAACTTAACGGCGGCACCTGCGCAAGGTT 1032
 QY 287 rTyrAspArgGluValThrProLeuIleGlySerProSerIleGlnProSerCysGly 306
 Db 1033 CTATTCGGGGCCCGACAGACCC-----GGGGCAGCGGAACCCCGCCCTTCTGT 1084

RESULT 10
 BI912313
 LOCUS
 DEFINITION 603068862F1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5217927 5',
 mRNA sequence.
 ACCESSION BI912313
 VERSION BI912313.1 GI:16176341
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 811)
 AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs@mail.nih.gov
 Tissue procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LMNL at:
<http://image.llnl.gov>
 Plate: LMNL1547 row: k column: 16
 High quality sequence stop: 805.

FEATURES
 source
 1..811
 Location/Qualifiers

/organism="Homo sapiens"
 /db.xref="taxon:9606"
 /clone="IMAGE:5217927"
 /clone_lib="NIH_MGC_118"
 /issue_type="Leukocyte"
 /lab_host="DH10B"
 /note="Vector: PCMV-SPO6; Site_1: NotI; Site_2: EcoRV
 (destroyed); RNA source leukocytes from anonymous pool of
 non-activated adult donors. Library is oligo-dT primed
 and directionally cloned (EcoRV site is destroyed upon
 cloning). Average insert size 1.7 kb, insert size range
 1.2-3.3 kb. Library is normalized and enriched for
 full-length clones and was constructed by C. Gruber
 (Invitrogen). Research Genetics tracking code 027. Note:
 this is a NIH_MGC Library."

BASE COUNT 185 a 206 c 300 g 120 t

ALIGNMENT SCORES:

Pred. No.: 5.34e-85 Length: 811
 Score: 953.50 Matches: 187
 Percent Similarity: 93.84% Conservative: 11
 Best Local Similarity: 88.63% Mismatches: 9
 Query Match: 44.31% Indels: 4

Alignment Scores:

Pred. No.: 3,26e-75 Length: 733
 Score: 855.00 Matches: 175
 Percent Similarity: 94.18% Conservative: 3
 Best Local Similarity: 92.59% Mismatches: 10
 Query Match: 39.73% Indels: 4
 DB: 14 Gaps: 1

US-09-068-377C-1 (1-415) x BQ207763 (1-733)

```

QY 228 ArgAnAlaLeuTTPVAlHIScysAnGlnLeuSerMetGlnCysValIlyAspAspGlu 247
    |||||
DB 733 CGCAATGCCCTGTGGTG-CACGTAAACAGAGCTCTCATGACAGTGTGCAAGAGCAGATAG 675
    |||||

QY 248 LeuTYrGluGluValAlaIrgLeuThrIleuGlyCysAspValGluGlyAspIleasnly 267
    |||||
DB 674 CTCTATGAGAGATCCGGCTGACACNTAGGGCTGTGATGTGAAGT-GATATCAATGC- 617
    |||||

QY 268 PheIleGlnSerIlySerThrGlyArgIupProPAlaProValProTyrGlnAsnTyr 287
    |||||
DB 616 TTCATCCAGTCCAAAGAGCAGTGGGAGAGAGCCCGAGCTCCCGTCCCTTATCAGAGACTAC 557
    |||||

QY 288 TyrAspArgIuValThrProleuIleGlySerProSerIleGlnProSerCysGlyVal 307
    |||||
DB 556 TACGACCGGAGAGTCCACCCACTGACTGAGACCCCTATGTCAGACCTCCTGTGGTGTG 497
    |||||

QY 308 IleIysArgPheSerGlyLeuLeuHISGlySerProIlyThrProSer---AlaPro 326
    |||||
DB 496 ATAAAGAGTCTCTCGGCTCTCTACATGAGATCCCAAGACACACCTTCTCAAGCTCT 437
    |||||

QY 327 AlaIaSerThrIrgIuThrLeuThrProThrProGluArgAsnGluLeuValTyrAlaSer 346
    |||||
DB 436 GCCGCTCCACAGAGACTGTGCTCCACCTCTGTGAGCAAGAAATTGCTTATGACATCC 377
    |||||

QY 347 IleGluValGlnAlaThrGlnGlyAsnLeuAsnSerSerIleGlnAspTyrArgAlaLeu 366
    |||||
DB 376 ATCGAAGTGCAGGACACCCAGGGAGAACTTAACTACAGACGCCAGGACTACCGGCGCTC 317
    |||||

QY 367 TyrAspTyrThrAlaGlnAsnSerAspGluLeuAspIleSerIleGlyAspIleLeuAla 386
    |||||
DB 316 TATGACTACACGAGACAGAAATTCTGATGAGCTGACATTTCTGGGGAGACATCTGCA 257
    |||||

QY 387 ValIleLeuGluGlyGluAspGlyTyrTrpThrValGluArgAsnGlyGlnArgGlyPhe 406
    |||||
DB 256 GTCATCTCTGAAAGGGAGGATGCTGTGAGCTGTGAAAGCAAGCAGACAGCTGGCTTT 197
    |||||

QY 407 ValProGlySerTyrLeuGluIlyLeu 415
    |||||
DB 196 GTCCCTGGTGTGACTTGGAGAACTC 170
    |||||

RESULT 15
BI910444 737 bp mRNA linear EST 16-OCT-2001
LOCUS 60306775F1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5216749 5',
DEFINITION mRNA sequence.
ACCESSION BI910444
VERSION BI910444.1 GI:16173833
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 737)
AUTHORS NIH-MGC http://mgs.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@pds-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:

http://image.lnl.gov
 Plate: LNL1544 row: j column: 14
 High quality sequence stop: 631.

FEATURES

source

```

1..737
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5216749"
/clone_id="NIH_MGC_118"
/tissue_type="Leukocyte"
/lab_host="DH10B"
/noise="Vector: pCMV-SPORT6; Site.1: NotI; Site.2: EcoRV
(destroyed); RNA source leukocytes from anonymous pool of
non-activated adult donors. Library is oligo-dT primed
and directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.7 kb, insert size range
1.2-3.3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(invitrogen). Research Genetics tracking code 027. Note:
this is a NIH-MGC library."
BASE COUNT 171 a 192 c 270 g 104 t
ORIGIN

```

Alignment Scores:

Pred. No.: 5.19e-73 Length: 737
 Score: 833.00 Matches: 173
 Percent Similarity: 89.86% Conservative: 13
 Best Local Similarity: 83.57% Mismatches: 16
 Query Match: 38.71% Indels: 5
 DB: 13 Gaps: 2

US-09-068-377C-1 (1-415) x BI910444 (1-737)

```

QY 1 MetMetAlaGlnLeuGlnPheArgAspAlaPheTrpCysArgAspPheThrAlaHisThr 20
    |||||
DB 117 ATGATGCCCGCAGCTGCAGTTCAAGATGCTTTTGGTGAGGAGACTTCACAGCCACAGC 176
    |||||

QY 21 GlyTyrIuValIleuLeuGlnArgLeuLeuAspIlyArgIyMetCysIlyAspValGlu 40
    |||||
DB 177 GCGTCAGAGTGTGCTGCGACGCGCTTCTGATGGCAGGAATGTGCAAGACATGGAG 236
    |||||

QY 41 GluLeuLeuArgGlnArgAlaGlnAlaGluGlnArgTyrGlySerGluLeuValGlnIle 60
    |||||
DB 237 GAGGTACTGAGCGACAGAGGCCCGAGCGGAGAGCGGTACGGGAGAGCTGTGTCAGATC 296
    |||||

QY 61 AlaArgIlyAlaGlyGlyGlnThrGluMetAsnSerLeuArgTyrSerPheAspSerLeu 80
    |||||
DB 297 GCACGGAAGCAGAGTGGCCAGACGAGATGACCTCCCTGAGGGCCTTGTGACTCCTTG 356
    |||||

QY 81 LysGlnGlnThrGluAsnValIlySerAlaHisIleGlnLeuAlaLeuAlaLeuArgGlu 100
    |||||
DB 357 AAGCAGCAAAATGGAAGATGTGGGACGCTCCACATCCAGTGGCCCTGACCTGCGTGGAG 416
    |||||

QY 101 GluLeuArgSerLeuGluGluPheArgGluArgGlnIlySerGluGlnArgIlySerTyrGlu 120
    |||||
DB 417 GAGCTGGGAGTCTCCAGAGAGTTCCTGAGAGGACAGAGAGCAGGAAGATGTGAG 476
    |||||

QY 121 AlaIleMetAspArgValGlnIlySerIlyLeuSerLeuTyrIlySerThrMetGluSer 140
    |||||
DB 477 GCCGTATGAGACCGGGCTCCAGAGACCAAGCTGTCCCTACAAAGAGCCATGGAGTCC 536
    |||||

QY 141 LysIlyAla-TyrAspGlnIlySerIlyArgAspAlaAspAspAlaGluGlnAlaPheGln 160
    |||||
DB 537 AAGAAAGCATTTAGACAGAGAGTGGCCGAGCCGGAGCAGCAGCGGAGGCTTCGAGC 596
    |||||

QY 160 rGValSerAlaAsnGlyHisGlnIlySerGlnValGluIlySerIlySerGlnAlaIlySer 179
    |||||
DB 597 GCATTAGCGGACAGGCGCCACCAAGACAGTGTGAGAGACATCAGAACACACAGCCAGGCC 656
    |||||

QY 179 Incys-LysGluSerAlaThrGluAlaGluArgValTyr--ArgGlnAsnIleGluGln 197
    |||||

```

Db 657 AGTGCACGGGACTCGGCCACCGAGGCGGATTCACAGACGACATTGCGCAG 716
QY 198 leuGluArgAlaArg 202
Db 717 CTGGAGAGGTCGCG 731

Search completed: August 3, 2003, 13:47:55
Job time : 1869 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: August 3, 2003, 12:08:41 ; Search time 67 Seconds
(without alignments)
1899.565 Million cell updates/sec

Title: US-09-068-377C-1

Perfect score: 2152
Sequence: 1 MAAQLQFRDAFMCDFTAHT.....WTVERNQGRGVFVCSYLEKL 415

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n model -DEV=rlh
-O/cgnt2_1/USFTO.spool/US09068377/runat_28072003_091317_15340/app_query.fasta.1.583
-DB=Issued_Patents_NA -OFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USRR=US09068377.ecgn.1.1.40.0runat_28072003_091317_15340 -NCPu=6 -ICPU=3
-NO.MMP -LARGEQUERY -NEG.SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: Issued_Patents_NA.*
2: /cgnt2_6/ptodata/1/ina/5A.COMB.seq.*
3: /cgnt2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgnt2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgnt2_6/ptodata/1/ina/PCPUS.COMB.seq.*
6: /cgnt2_6/ptodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2152	100.0	2100	3	US-08-938-830-2
2	2152	100.0	2100	3	US-09-020-222-2
3	1913.5	88.9	1858	4	US-09-006-428A-16
4	1748	81.2	1803	4	US-09-006-428A-18
5	1714.5	79.7	1613	3	US-08-938-830-28
6	260.5	12.3	4091	4	US-08-630-915A-33
7	220.5	10.2	1133	4	US-08-630-915A-35
8	181	8.4	3345	4	US-08-630-915A-29
9	179.5	8.3	1302	4	US-08-630-915A-17
10	176.5	8.2	1531	4	US-08-630-915A-21
11	166	7.7	1867	4	US-08-630-915A-13
12	154.5	7.2	549	4	US-08-630-915A-199

13	153	7.1	1687	4	US-08-630-915A-191	Sequence 191, App
14	150.5	7.0	1925	1	US-08-435-454-3	Sequence 3, Appli
15	150.5	7.0	1925	2	US-08-919-145-5	Sequence 5, Appli
16	150.5	7.0	1925	3	US-09-344-889-5	Sequence 5, Appli
17	150.5	6.9	1925	5	PCT-US96-06231A-3	Sequence 3, Appli
18	149.5	6.9	1583	4	US-09-370-838-224	Sequence 224, App
19	149.5	6.9	1925	1	US-08-652-972A-3	Sequence 3, Appli
20	149.5	6.9	1925	1	US-08-870-126-3	Sequence 3, Appli
21	149.5	6.9	1925	4	US-09-445-247-3	Sequence 3, Appli
22	148.5	6.9	2080	2	US-08-878-563A-2	Sequence 2, Appli
23	148.5	6.9	2080	4	US-09-270-117-2	Sequence 2, Appli
24	147.5	6.9	6306	1	US-08-466-390-3	Sequence 3, Appli
25	147.5	6.9	6306	1	US-08-470-950-3	Sequence 3, Appli
26	147.5	6.9	6306	1	US-08-467-781-3	Sequence 3, Appli
27	147.5	6.9	6306	1	US-08-195-487-3	Sequence 3, Appli
28	147.5	6.9	6306	2	US-08-483-924-3	Sequence 3, Appli
29	147.5	6.9	6306	5	PCT-US93-06160-3	Sequence 3, Appli
30	147	6.8	3846	2	US-08-691-814B-3	Sequence 3, Appli
31	146.5	6.8	2873	4	US-08-630-915A-193	Sequence 193, App
32	144	6.7	6921	4	US-09-643-597-117	Sequence 117, App
33	143	6.6	1199	4	US-08-630-915A-15	Sequence 15, Appl
34	139.5	6.5	1734	4	US-08-630-915A-23	Sequence 23, Appl
35	137	6.4	1659	1	US-08-475-894-1	Sequence 1, Appli
36	137	6.4	1659	1	US-08-484-710-1	Sequence 1, Appli
37	137	6.4	1659	2	US-08-484-709-1	Sequence 1, Appli
38	137	6.4	1659	4	US-08-474-697-1	Sequence 1, Appli
39	137	6.4	2608	4	US-09-154-750A-75	Sequence 75, Appl
40	137	6.4	3143	4	US-08-671-354-1	Sequence 1, Appli
41	136.5	6.3	1458	2	US-08-942-423-68	Sequence 68, Appl
42	136.5	6.3	2003	2	US-08-942-423-1	Sequence 1, Appli
43	136	6.3	1045	1	US-08-475-894-5	Sequence 5, Appli
44	136	6.3	1045	1	US-08-484-710-5	Sequence 5, Appli
45	136	6.3	1045	2	US-08-484-709-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-08-938-830-2
Sequence 2, Application US/08938830
Patent No. 6040437
GENERAL INFORMATION:
APPLICANT: Lasky, Laurence A.
APPLICANT: Dowbenko, Donald J.
TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESS: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938, 830
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/798419
FILING DATE: 07-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33, 055
REFERENCE/DOCKET NUMBER: P1066P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-3216
TELEFAX: 650/952-9881

; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2100 base pairs
 ; TYPE: Nucleic Acid
 ; STRANDEDNESS: Single
 ; TOPOLOGY: Linear
 ; US-08-938-830-2

Alignment Scores:

Pred. No.:	1.28e-216	Length:	2100
Score:	2152.00	Matches:	415
Best Local Similarity:	100.00%	Conservative:	0
Query Match:	100.00%	Mismatches:	0
DB:	3	Indels:	0
		Gaps:	0

US-09-068-377c-1 (1-415) x US-08-938-830-2 (1-2100)

```

QY      1 MetMetAlaGlnLeuGlnPheArGAspAlaPheTrpCyArGAspPheThrAlaHisThr 20
DB      682 ATGATGGCCACCTGAGTTCGAGATGCTTGTGTCAGAGACTTCACGGCCACACA 741

QY      21 GIYTrGluValLeuLeuGlnArGLeuAspGlyArGLeuMetCysIAspValGlu 40
DB      742 GGGTATGAGCTCTACTGTCAGAGGCTGTCGAGAGGAGAGATGTGCAAGATGTGAG 801

QY      41 GluLeuLeuArGlnArGAlaGlnAlaGluArGTrpGlyLysGluLeuValGlnIle 60
DB      802 GAGCTGCTCAGACAGAGGGCCAGGGGAGAGAGGTACGGAGAGAGGTGTGTCAGATT 861

QY      61 AlaArGlyAlaGlyGlyGlnThrGluMetAsnSerLeuArGTrpSerPheAspSerLeu 80
DB      862 GCACCAAGGCTGGGGCCAGACAGAGATGATTCCTGAGACCTCTTTCACCTCCG 921

QY      81 LysGlnGlnThrGluAsnValGlySerAlaHisIleGlnLeuAlaLeuArGlu 100
DB      922 AAGCAGCAACAGCAAAATGTGGGCACTGCACATCCAGCTGGCCCTGGCTGGTGAG 981

QY      101 GluLeuArGSerLeuGluGluPheArGLeuArGlnGlyGlnArGLeuLysTrpGlu 120
DB      982 GAGCTGCGGAGCTTGAGAGAGTTCGAGAGAGACAGAAAGACAGCAGAGAGATGAG 1041

QY      121 AlaIleMetAspArGValGlnLysSerLysLeuSerLeuTrpLysLysThrMetGluSer 140
DB      1042 GCCATCATGACCGGTCAGAAAGAGCAAGTTGCTGCTACAGAAAGACCATGAGATGCC 1101

QY      141 LysLysAlaTrpArGLeuGlnLysCysArGAspAlaAspAspAlaGluGlnAlaPheGluArG 160
DB      1102 AAGAAGGCATATGACACAGAAAGTGCAGGATGATGATGCTGACAGAGCCCTTCAGAGCT 1161

QY      161 ValSerAlaAsnGlyHisGlnLysGlnValGluLysSerGlnAsnLysAlaLysGlnCys 180
DB      1162 GTGAGTCCCATGCGCCACACAGAGCAAGTAGAAAAGACCCAGAACCAAGCAGAGTCC 1221

QY      181 LysGlnSerAlaThrGlnAlaGlnArGValTrpArGlnAsnIleGluGlnLeuGluArG 200
DB      1222 AAGAGTCTAGCCACAGAGCAAGAAAGAGTACAGGCAAAATATGAAACAACACTGGAGAGA 1281

QY      201 AlaArGTrpGlnTrpGlnGlnGlnHisArGTrpThrCysGlnValaPheGlnLeuGlnGlu 220
DB      1282 GCGAGGACCGAGTGGGAGCAGAGACACCGGACTACCTGTGAGGCCCTTCACAGTTCAGAGAG 1341

QY      221 PheAspArGLeuThrIleLeuArGAsnAlaLeuTrpValHisCysAsnGlnLeuSerMet 240
DB      1342 TTTCACCGGCTCACCATCTCCGCAATGGCTGTGGGTGACCTGTAAACCACTCTCCATG 1401

QY      241 GluCysValLysAspAspGluLeuTrpGluValaArGLeuThrLeuGluGlyCysAsp 260
DB      1402 CAGTGTCTCAAGGATGAGACTTATGAGAAATGGCGGACCTTGAGGGCTGTGAT 1461

QY      261 ValGluGlyAspIleAsnGlyPheIleGlnSerLysSerTrpGlyArGTrpGluProProla 280
DB      1462 GTGGAAGCTGACATCATGCTTCATCCAGAGCACTGGCAGAGAGAGAGAGAGAGAGAGT 1521
  
```

```

QY      281 ProValProGlyGlnAsnTrpTrpAspArGLeuValThrProLeuIleGlySerProSer 300
DB      1522 CCGGTCCCTTATCAGAACTACTATGACAGAGGAGTGACCCACTGATTTGGAGCCCTTAC 1581

QY      301 IleGlnProSerCysGlyValIleLysArGPhaSerGlyLeuLeuHisGlySerProLys 320
DB      1582 ATCCACCCCTCCCGCGGTGTGATMAAGAGTTCTCTGGGCTGCTACATGGAAGTCCCAAG 1641

QY      321 ThrThrProSerAlaProAlaAsnThrGluTrpLeuThrProThrProGluArGAsn 340
DB      1642 ACCACACTTCTGCTCTCTCTCTCTCCACAGAGACTGACTCCACCCCTGAGGGAGAT 1701

QY      341 GluLeuValTrpAlaSerIleGluValaGlnAlaThrGlnGlyAsnLeuAsnSerAla 360
DB      1702 GAGTTGTCTACGATCTTCATCGAAGTGCAGGGAGCCAGGAGAAACCTTAACATCATCAGCC 1761

QY      361 GlnAspTrpArGAlaLeuTrpAspTrpThrAlaGlnAsnSerAspGluLeuAspIleSer 380
DB      1762 CAGGACTACCGGGGCACTCTACGACTACACTGCACAGAAATTTGATGAGCTGAGACATTTCC 1821

QY      381 AlaGlyAspIleLeuAlaValIleLeuGluGlyLysAspGlyTrpTrpThrValGluArG 400
DB      1822 GCGGAGAGCATCCTGCGGCTCATCTCGAAGGAGGAGAGATGCTGTGAGCTGTGAGCGG 1881

QY      401 AsnGlyGlnArGLeuPheValProGlySerTrpLeuGlyLysLeu 415
DB      1882 AACGACAACTGGCTTTGTCTTGTGCTGAGTCTTGTGAGAAAGCTC 1926
  
```

RESULT 2
 US-09-020-222-2
 ; Sequence 2, Application us/09020222
 ; Patent No. 6111073

GENERAL INFORMATION:
 APPLICANT: Lasky, Laurence A.
 TITLE OF INVENTION: Tyrosine phosphorylated cleavage
 TITLE OF INVENTION: Furrow-Associated Proteins (FSTPSPs)
 NUMBER OF SEQUENCES: 27
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genentech, Inc.
 STREET: 1 DNA Way
 CITY: South San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94080

COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5 inch, 1.44 MB floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Winpatin (Genentech)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/020, 222
 FILING DATE: 06-Feb-1998

CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/798419
 FILING DATE: 02/07/1997

ATTORNEY/AGENT INFORMATION:
 NAME: Dreyer, Ginger R.
 REGISTRATION NUMBER: 33,055
 REFERENCE/DOCKET NUMBER: P1066r1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 650/225-3216
 TELEFAX: 650/952-9881

INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2100 base pairs
 TYPE: Nucleic Acid
 STRANDEDNESS: Single
 TOPOLOGY: Linear
 US-09-020-222-2

Alignment Scores:

Db 680 AACCAACAAATGAGATGTGGCAGCTCACATCCAGCTGGCCCTGGACCTGGCTGAG 739
QY 101 GATleuArSerleuGluGluPheArgLuarGlnLysGluGlnArgLysLysTYRGu 120
Db 740 GAGCTCGGAGTCTCCAGAGATTCTGTGAGAGCAGAGAGAGAGAGAGAGATATGAG 799
QY 121 AlaIleMetAspArgValGlnLysSerLysLeuSerLeuTYRlysTYRMetGluSer 140
Db 800 GCCGTATGAGACCGGGTCCAGAGAGAACCTGCTGCTACAGAGAGAGGACCATGAGATCC 859
QY 141 LysLysAlaTYRAspGlnLysCysArgAspAlaAspAspAlaGluGlnAlaPheGluArg 160
Db 860 AAGAAACATACAGAGAGAGTCCCGGAGCGGAGCGAGCGAGCGAGCGCTTCGAGCGC 919
QY 161 ValSerAlaAsnGlnLysGlnLysGlnValGlnLysSerLysAsnLysAlaLysGlnCys 180
Db 920 ATTAGCGCCACAGCGCCACCAAGAGCAGTGGAGAGAGTGAAGAACAAAGCCAGCAGTGC 979
QY 181 LysGluSerAlaThrGlnAlaGluArgValTYRArgGlnAsnIleGluGlnLeuArg 200-
Db 980 AAGACTCGGCGCCAGCGAGCGAGCGGATACAGGAGAGCATGGCGCAGTGGAGAG 1039
QY 201 AlaArgThrGluTYRProGluGlnLysArgTYRThrCysGluAlaPheGlnLeuGln 220
Db 1040 GTCCGGGCTGAGTGGAGAGAGAGACCGGACCGGACCGCTTGAGGCTTCAGCTGCAAGAG 1099
QY 221 PheAspArgLeuThrIleLeuArgAsnAlaLeuTYRValHisCysAsnGlnLeuSerMet 240
Db 1100 TTGGACCGGCTGACATTCTCCGCAAGCCCTGGGTGGCAGCAGCAACGCTCTCATG 1159
QY 241 GlnCysValLysAspAspGluLeuTYRArgLysGlnValArgLeuThrLeuGlnLysAsp 260
Db 1160 CAGTGTCTCAAGATGATGAGTCTACAGAGAGAGTGGGGCTGACGCTGGAAGCTGAGC 1219
QY 261 ValGluGlnAspLysLeuGlnPheIleGlnSerLysSerThrGlnArgGluProProAla 280
Db 1220 ATAGACCGCAGCATTCAGATTTCACAGGCAAGGCGGAGCGGACAGGCGCCGCT 1279
QY 281 ProValProTYRAsnTYRAspArgLysValThrProLeuIleGlnSerProSer 300
Db 1280 CCGGTCCTGACAGCACTATTCAGATCGGAGGTGACCGCGCTGACACACACCGCTGGC 1339
QY 301 IleGlnProSerCysGlnValIleLysArgPheSerGlyLeuLeuHisGlySerProLys 320
Db 1340 ATACACCGCTCCCGCGCATGATMAAGAGGTCTCTGAGCTGACAGGAGATCCCAAG 1399
QY 321 ThrThrPro--SerAlaProAlaIleSerThrGluThrLeuThrProThrProGluArg 339
Db 1400 ACCACTTCTGGCAGCTCTGCTGGCTCCACAGAGACCTGACCCCGCCGAGCGG 1459
QY 340 AsnGluLeuValTYRAlaSerIleGluValGlnAlaThrGlnLysAsnLeuAsnSerSer 359
Db 1460 AATGAGGGTGTCTACACAGCATTCGACAGAGATACAGGGAACCCGGGCTTCACCA 1519
QY 360 AlaGlnAspTYRArgAlaLeuTYRAspTYRThrAlaGlnAsnSerAspGluLeuAspIle 379
Db 1520 GCCCAGAGATACCGCGCTCTACGATTAATACAGCCAGAACCCAGATGAGCTGGAGCTG 1579
QY 380 SerAlaGlnAspIleLeuAlaValIleLeuGlnLysGlnAspGlyTYRThrProThrValGlu 399
Db 1580 TCCGGCGGAGACATCTTGAGTGATCTCGAAGGAGGAGATGCTGTGTGACTGTGGAG 1639
QY 400 ArgAsnGlyGlnArgLysPheValProGlySerTYRLeuGlnLysLeu 415
Db 1640 AGCAAGCGGAGCGTGGCTTCGCTGCTTCTACTGTGAGAACCTT 1687

RESULT 4
US-09-006-428A-18

Sequence 18, Application us/09006428A

Patent No. 644439

GENERAL INFORMATION:

APPLICANT: JUNG LI
Kazuhisa Nishizawa

APPLICANT: Menglan An
APPLICANT: Ellis L. Reinherz
TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF A
FILE OF INVENTION: cdc15-LIKE ADAPTOR PROTEIN (CD2BP1)
FILE REFERENCE: 1062.1020-000
CURRENT APPLICATION NUMBER: US/09/006.428A
NUMBER OF SEQ ID NOS: 28
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 18
LENGTH: 1803
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (440)...(1630)
US-09-006-428A-18
Alignment Scores:
Pred. No.: 3,14e-174 Length: 1803
Score: 1748.00 Matches: 341
Percent Similarity: 86.78% Conservative: 20
Best Local Similarity: 81.97% Mismatches: 35
Query Match: 81.23% Indels: 20
Gaps: 2
US-09-068-377c-1 (1-415) x US-09-006-428A-18 (1-1803)
QY 1 MetMetAlaGlnLeuGlnPheArgAspAlaPheTYRAspArgAspPheThrAlaHisThr 20
Db 440 ATGATGCCCGCAGCTGCGCTCAAGATGCTTTGGTGGCAGGAGCTTCACAGCCACAGC 499
QY 21 GlnTYRAlaValLeuLeuGlnArgLeuLeuAspGlyValArgLysMetCysLysAspValGlu 40
Db 500 GCGTACAGAGTCTGCTCTGCGACCGGCTTCTGATGCGCAGAGATGTGCAAGACATGGAG 559
QY 41 GlnLeuLeuArgGlnArgAlaGlnAlaGlnLysGlnLysGlnLysGlnLysGlnLys 60
Db 560 GAGCTACTGAGCAGAGAGGCGCCAGCGGAGCGAGCGGATACGGAAGAGAGTGTGAGATC 619
QY 61 AlaArgLysAlaGlnLysGlnThrGlnMetLysSerLeuArgThrPheAspSerLeu 80
Db 620 GCAGGAGGAGCAGTGGCGCCAGACGAGATCACTCCCTGGGCTCTTGTGACTCTTG 679
QY 81 LysGlnGlnThrGlnLysAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaLeuArgGlu 100
Db 680 AAGCAGGAATGAGAGATGTGGGAGCAGTCAACATGCAGTGGCCCTGACCGCGGTGAG 739
QY 101 GlnLeuArgSerLeuGlnLysPheArgGlnArgGlnLysGlnLysGlnLysLysTYRGu 120
Db 740 GAGCTGGGAGTCTCGAGAGTTTCGTGAAGCGCAGAGAGAGAGAGAGATATGAG 799
QY 121 AlaIleMetAspArgValGlnLysSerLysLeuSerLeuTYRlysTYRMetGluSer 140
Db 800 GCCGTATGAGACCGGGTCCAGAGAGCAACTCTGCTGTACAGAGAGAGGACCATGAGATCC 859
QY 141 LysLysAlaTYRAspGlnLysCysArgAspAlaAspAspAlaGluGlnAlaPheGluArg 160
Db 860 AAGAAACATACAGAGAGAGTCCCGGAGCGGAGCGAGCGAGCGAGCGCTTCGAGCGC 919
QY 161 ValSerAlaAsnGlnLysGlnLysGlnValGlnLysSerLysAsnLysAlaLysGlnCys 180
Db 920 ATTAGCGCCACAGCGCCACCAAGAGCAGTGGAGAGAGTGAAGAACAAAGCCAGCAGTGC 979
QY 181 LysGluSerAlaThrGlnAlaGluArgValTYRArgGlnAsnIleGluGlnLeuGluArg 200
Db 980 AAGACTCGGCGCCAGCGAGCGAGCGGATACAGGAGAGCATTCGCGCAGTGGAGAG 1039
QY 201 AlaArgThrGluTYRProGluGlnLysArgTYRThrCysGluAlaPheGlnLeuGln 220
Db 1040 GTCCGGGCTGAGTGGAGAGAGAGACCGGACCGGACCGCTTGAGGCTTCAGCTGCAAGAG 1099
QY 221 PheAspArgLeuThrIleLeuArgAsnAlaLeuTYRValHisCysAsnGlnLeuSerMet 240

DB 1100 TTGACGGCTGACATTCTCCGACAGCGCCTGTGGTGACAGCAGACAGCTCTCATG 1159
QY 241 GACysValLysAspAspGluLeuTyrgluGluValArgLeuThrLeuGluGlyCysAsp 260
DB 1160 CAGTGTCAAGAGTATGAGCTCTACGAGAGAGTCCGGCTGACGTGAGAGCTGCACG 1219
QY 261 ValGluGlyAspLeuSnglPheLeuGlnserLeuSerThrGlyArgGluProProAla 280
DB 1220 ATAGAGCGCCAGACAGATTTCATCCAGGCGCAAGAGCGGCGACAGAGCGCCCGGT 1279
QY 281 ProValProTyrglnAsnTyrrTyraAspArgGluValThrProLeuIleGlySerProser 300
DB 1280 GAGGTCCGGCTGCGGAC----- 1297
QY 301 IleGlnProSerCysGlyValIleLysArgPheSerGlyLeuLeuHISGlySerProLys 320
DB 1298 -----AGCCGAGCCTCTAGGTCTCTGAGCTGTGACGAGGAACTCCCAAG 1342
QY 321 ThrThrPro---SerAlaProAlaIleSerThrGluThrLeuThrProThrProGluArg 339
DB 1343 ACCACTTCGTTGGCAGCTTCTGCTGCTCCACAGAGACCTGACCCCGGAGCGG 1402
QY 340 AsnGluLeuValTyrglnAsnIleGluValGlnAlaThrGlnLysAsnLeuAsnSerSer 359
DB 1403 AATGAGGCTGTCTACACAGCCATCCGACGTGACAGAGATACAGGAAACCGGCTCACCA 1462
QY 360 AlaGlnAspTyrrArgAlaLeuTyraAspTyrrThrAlaGlnAsnSerAspGluAspIle 379
DB 1463 GCCCAGAGTACCGGCGCTCTCTGATTTATACAGCCAGAACCCAGATGAGTGAGACTG 1522
QY 380 SerAlaGlyAspIleLeuAlaValIleLeuGluGlyGluAspGlyThrPrrThrValGlu 399
DB 1523 TCCGCGGAGACATCTCTGAGGTATCTGGAAAGGAGATGCTGTGTGACTGTGAG 1582
QY 400 ArgAsnGlyGlnArgGlyPheValProGlySerTyrrLeuGluLysIleu 415
DB 1583 AGGAGCGGACAGCGTGTCTCTGCTCTCTACTGTGAGAAAGCTT 1630

RESULT 5
US-08-938-830-28
Sequence 28, Application US/08938830
Patent No. 6040437
GENERAL INFORMATION:
APPLICANT: Lasky, Laurence A.
TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938, 830
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/798419
FILING DATE: 07-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: P1066P1
TELECOMMUNICATION INFORMATION:

TELEPHONE: 650/225-3216
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1613 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
US-08-938-830-28

Alignment Scores:
Pred. No.: 8,82e-171 Length: 1613
Score: 1714.50 Matches: 338
Percent Similarity: 81.80% Conservative: 17
Best Local Similarity: 77.88% Mismatches: 26
Query Match: 79.67% Indels: 53
DB: Gaps: 3

US-09-068-377c-1 (1-415) x US-08-938-830-28 (1-1613)

QY 1 MetMetAlaGlnLeuGlnPheArgAspAlaPheTrpCysArgAspPheThrAlaHisThr 20
DB 344 ATGATGCCCGCCAGCTGTGCAATCAAGATCCCTTTGTGAGGAGGACTTCACAGCCACAGC 403
QY 21 GLYTyrGluValLeuLeuGlnArgLeuLeuAspGlyArgLysMetCysLysAspValGlu 40
DB 404 GGTACAGAGTGTCTCTGCGACGGCTTCTGATGGCAGAGATGTCCANAGCATGTGAG 463
QY 41 GlnLeuLeuArgGlnArgAlaGlnAlaGluGluArgTyrGlyLysGluLeuValGlnIle 60
DB 464 GACCTACTGAGCAGAGAGGCGCCAGCGAGAGCGGTACGGAAGAGCTGTGTGACATC 523
QY 61 AlaArgLysAlaGlyGlyGlnThrGlnMetAsnSerLeuArgThrPheAspSerLeu 80
DB 524 GCACGGAGGACAGGTGGCCAGACGAGATCACTCCCTGAGGGCTCTTTGACTCTTG 583
QY 81 LysGlnGlnThrGlnAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaLeuArgGlu 100
DB 584 AAGCAGCAATGTGAGATGTGGCGAGCTCACATCCAGCTGGCCCTGACCTCCGTG 643
QY 101 GlnLeuArgSerLeuGlnGluPheArgGluArgGlnLysGlnLysGlnArgLysLys----- 118
DB 644 GAGCTGGGAGTCTCCGAGAGTTTCTGAGAGGAGCAGAGAGGAGGAGGAGCATG 703
QY 119 -----TyrGluAla 121
DB 704 GCTGTCCCGACAGACAGTACTGATGGAAGTGAAGTCCCATCATCAGGAGTATGAGGCC 763
QY 122 IleMetAspArgValGlnLysSerLysLeuSerLeuTyrrLysLysThrMetGluSerLys 141
DB 764 GTCATGTGACCGGGGTCCGAAAGAGCAAGCTGTCTGTACAAAGAGGCATGTGAGTCCAG 823
QY 142 LysAlaTyraSpGlnLysCysArgAspAlaAspAlaGluGlnAlaPheGluArgVal 161
DB 824 AAGCATATCAGCAGAGAGTCCCGGAGCGGAGCAGCAGCGAGCGCTTCGAGCGCAT 883
QY 162 SerAlaAsnGlnLysGlnLysGlnValGlnLysSerLysAsnLysAlaLysGlnLysLys 181
DB 884 AGCGCAACGCGCCAGGAGAGAGTGAAGAGTGAAGCAAAAGCAGGACAGTGGAG 943
QY 182 GluSerAlaThrGlnAlaGluArgValTyraArgGlnAsnIleGlnGlnLeuGluArgAla 201
DB 944 GACTCGGCCACCGACGAGAGCGGTATACAGGAGAGCATTTGGCGAGCTGTGAGAAAGTTC 1003
QY 202 ArgThrGluTrpGluGlnGluHisArgThrThrCysGlnAlaPheGlnLeuGlnLysPhe 221
DB 1004 CGGCTGTAGTGGGAGCAGAGCAGCCGACACCTGTAGGCTTTGACGTGCAAGATT 1063
QY 222 AspArgLeuThrIleLeuArgAsnAlaLeuThrValHisCysAsnGlnLeuSerMetGln 241
DB 1064 GACCGCTGACCATTTCTCCGAAAGCGCTGTGGGTGACAGTAAACAGCTCTCCATGACG 1123
QY 242 CysValLysAspAspGluLeuTyrgluGluValArgLeuThrLeuGluGlyCysAspVal 261

Db	1124	TGTGTCAAGATGATGAGCTCTACGAGGAATGCGCGCTGACGCGCTGGAAAGCTGCACGATTA	1183
QY	262	GIuGIyAspIleasnGIyPheIleGIserLysserThrGIyArgGIuProProAlaPro	281
Db	1184	GACCCCGCATCGACAGTTTCATCCAGGCCAAGAGACAGGGCAGACAGACCCCC	1237
QY	282	ValProTIyGIuAsnTIyTyrAspArgGIuValThrProLeuIleGIySerProSerIle	301
Db	1237	-----	1237
QY	302	GIuProSerCysGIyValIleLysArgPheSerGIyLeuHISGIySerProLysThr	321
Db	1238	-----AGGTTCTCTGACTGCTGCACGGAGTCCCAAGACC	1273
QY	322	ThrProSerAlaProAlaAlaSerThrGIuThrLeuThrProThrProGIuArgAsnGIu	341
Db	1274	ACTTCGTCAGCTTCGTGCTGCCTCCACAGACAGACCCGTGACCCCGCAGCGGAATGAG	1333
QY	342	LeuValTIyAlaSerIleGIuValGIuAlaThrGIuAsnLeuAsnSerSerAlaGIu	361
Db	1334	GGTTCATACACAGCATCGCAGTGCAGAGATACAGGGAACCCGGCTTCACAGCCGAG	1393
QY	362	AspTIyArgAlaLeuTIyTyrAspTIyThrAlaGIuAsnSerAspGIuLeuAspIleSerAla	381
Db	1394	GACATCCGGGGCGCTCTGACGATTATACACGCCAGAACCCAGATGACCTGACCTGCGCG	1453
QY	382	GIyAspIleLeuAlaValIleLeuGIuGIyLysAspGIyTyrPThrValGIuArgAsn	401
Db	1454	GGAGAC-----ATCCTGGAAGGGAGGATGGCTGTGGACTGTGAGAGGAC	1501
QY	402	GIyGIuArgGIyPheValProGIySerTIyLeuGIuLysLeu	415
Db	1502	GGGAGCGGTGGCTTCGCCCTGGTTCTCTACCTGGAGAAAGCTT	1543

RESULT 6
 US-08-630-915A-33
 Sequence 33, Application US/08630915A
 Patent No. 6309820
 GENERAL INFORMATION:
 APPLICANT: SPARKS, Andrew B.
 APPLICANT: HOFFMAN, No. 6309820h
 APPLICANT: KAY, Brian K.
 APPLICANT: FOLKES, Dana M.
 APPLICANT: MCCONNELL, Stephen J.
 TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
 TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF
 TITLE OF INVENTION: USING SAME
 NUMBER OF SEQUENCES: 227
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds LLP
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10036-2711
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/630,915A
 FILING DATE: 03-APR-1996
 CLASSIFICATION: 536
 ATTORNEY/AGENT INFORMATION:
 NAME: Mistrock, S. Leslie
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 1101-174
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9090
 TELEFAX: (212) 686-8864/9741
 TELEX: 66141 PENNIE

```

; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 4091 bases
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA
;   US-08-630-915A-33

Alignment Scores:
Pred. No.: 9,44e-18      Length: 4091
Score: 265.50           Matches: 100
Percent Similarity: 39.07%      Conservative: 84
Best Local Similarity: 21.23%   Mismatches: 192
Query Match: 12.34%            Indels: 95
DB: 4                      Gaps: 16

US-09-068-377C-1 (1-415) x US-08-630-915A-33 (1-4091)

OY      3  AlagInleuGInDhPaarGaSPaLaPheTrCYsArGaSPaRheThAlaHisThGly- 21
Db      221  GCCTCAGAGGAGATCATCAGATGAGCTTCTG-----GAGGTGGGGAAC 265
OY      22  TyrGluValLeuLeuGInAlnArgLeuLeuSerGlyAlaGlySMetCysLysAspValGlu 41
Db      263  TACAAAGGAGCGGTGAAGCCGATCGACGATGGGGACCCGCTGTGCACAGCCTTAGAGC 322
OY      42  LeuLeuArgGInAlnArgAlaGInAlaGluAlaGlyTyrGlyLysGluLeuValGlnLea 61
Db      323  TGCCTGCAGAGAGCGCCGCAAGATCGAAGAGCATGCGGACGACAGCTACCGAGTGGGCC 382
OY      62  ArgLys-----AlaGlyGInThrGluMetLanSerLeuArgThr 75
Db      383  AAGGGCTGGCGCCAGCTCATCGAAGAGGTCTCTGAG-----TAGGAGCCTTGAGCGG 436
OY      76  SerPheAspSerLeuLysGInGInThrGluAsnValGlySerAlaHisIleGInLeuAla 95
Db      437  GCGGGGGGCCCATGTGATGACAGAAAGCATTAAGTCTACAGCAGCTGCACACAGAGTGAAG 496
OY      96  LeuAlaLeuArgGluGluLeuLeuArgSerLeuGluGluGluPheArgGluArgGlnGlu 114
Db      497  AACAGCTGCTGATGATGAG-----GACGTGGGAAGATGCAAGAACTGGCGAAGAGATGCC 550
OY      115 -----GlnArgLysLysTyrGluAlaIleMetAsp 124
Db      551  TATTCACAGACAGATCATGTGGGTGGCTTCAAGACAGAGAAAGAGCGCGAGATGCGTTCGA 610
OY      125  ArgValGInLysSerLysLeuSerLeuTyrTylsLysThrMetGluSerLysAlaTyr 144
Db      611  AAGGCCACAGAAAGCCTGGGCTAAAGATGAAGAGAGCTAAGAGCGGCGCCAAAGAGCCTAT 670
OY      145  AspGlnLysSerArgSPaLaSPaSPaLa-----GluGInAlaPheGluArgValSer 162
Db      671  CACTTGAGCTGTGTAAGGAGAAAGCGTGGCCATGACCCGCGAGATGAACAGTAAACAGAG 730
OY      163  AlaAsnGlyHisGInLysGInValGluLysSerGlnAsnLysAlaLysGlnCysLysGlu 182
Db      731  CACTCGGTCAACCCCTGAACAGACAGAAAGAACTGTGGACAAAGTGGACAAATGACAGCAG 790
OY      183  SerAlaThrGluAlaGluArgValTyrArgGlnAsnIleGluGInLeuGluArgAlaArg 202
Db      791  GATGTCAAAAGACATCAGAGGAAGATGATGAAGAGCTCTCGAAGATGTGGGCAAGACCACA 850
OY      203  ThrGluThrGluGlnGlnHisArgThrThrCysGluAlaPheGlnGlnGluPheAsp 222
Db      851  CCACAGATACATGAGGCGCATGAGCAGAGGTGTTTGGACGCTCCAGCATTTGAGAGAAC 910
OY      223  ArgLeuThrIleLeuArg----- 228
Db      911  CGGCTGTCTCTCTGAAGAGACTCCTGCTGATATCAAAAGCGCATTCAACTTACGCGAG 970
OY      229  AsnAlaLeuThrAlaIscysAsnGInLeuSerMetGlnCysValLysAspArgGluLeu 248

```

Db 971 AACAGACGTACATGCAT-----GTC 991
Qy 249 TyrgluValArgLeuThrLeuGluGlycysAspValGluGlyAspIleAsnGlyPhe 268
Db 992 TACCGAAGATCGAGCAGCAGCCATCCGGGGGGGCGCCGACCGACCTCAGGTGGTTC 1051
Qy 269 IleGlnSerLysSerThrGlyArgGluProAlaProValProTyrGlnAsnTyr 288
Db 1052 CGCAGACCACTGGCCCGGGGATG-----CCATGACACTGGCCCGACCTGCAGAGTGC 1105
Qy 289 AspArgGluValThrProLeuIleGlySerProSerIleGlnProSerGlyValIle 308
Db 1106 AACCCAGACCTCCCGCACACCACTCCCAAGAGAAAGAACCCCT----- 1150
Qy 309 LysArgPheSerGlyLeuLeuHisGlySerProLysThrThrProSerAlaProAla 328
Db 1151 AAGAGCGCAGAGGG-----GCCACCTGAGCAATGCCACTGG 1189
Qy 329 SerThrGluThrLeuThrProThrProGluArgAsnGluLeu----- 342
Db 1190 GCTGAGATCCACATCCAGCAGCTGGGAGCCGCTGACAGTGTAGCAGCTATGACCGAGGC 1249
Qy 343 --ValTyrAlaSerIleGluValGlnAlaThrGlnGlyAsn----- 355
Db 1250 CAACATATGCCACCGAGTGGTCAAGCATGAGACGGAACCCCTCGGGGCAATGAG 1309
Qy 356 LeuAsnSerSerAlaGlnAspTyr-----ArgAlaLeu 366
Db 1310 GCCAATGCTGGCGCAACCCCTTCAGAGATGATGCCAAGGAGCTTCGTACGGGCACCTC 1369
Qy 367 TyrAspTyrThrAlaGlnAsnSerAspIleLeuAspIleSerAlaGlyAspIleLeuAla 386
Db 1370 TATGACTACGACGCGTACAGCAGCATGAGCTCAGCTTCACAGCGCGAGATAGCTCAC 1429
Qy 387 ValIleLeuGluGlyGluAspGlyTyrThrThrValGluArg-----AsnGlyGlnArg 404
Db 1430 AAGTCGAGAGAGGAAGAACAGAGGCTGTGCCCGCGCGCTGACAGCGCAGCATGCTG 1489
Qy 405 GlyPheValProGlySerTyrLeuGluLysLeu 415
Db 1490 GCCCTATCTCTGCCAAGCTAGCTGAGGCTATA 1522

RESULT 7
US-08-630-915A-35
Sequence 35, Application US/08630915A
Patent No. 6309820

GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
APPLICANT: HOFFMAN, No. 6309820H
APPLICANT: KAY, Brian K.
APPLICANT: FOWLES, Dana M.
APPLICANT: MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Penne & Edmunds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:

NAME: Mirock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 1133 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-630-915A-35

Alignment Scores:
Pred. No.: 7,04e-14 Length: 1133
Score: 220.50 Matches: 90
Percent Similarity: 38.97% Conservative: 62
Best Local Similarity: 23.08% Mismatches: 149
Query Match: 10.25% Indels: 90
Gaps: 12

US-09-068-377C-1 (1-415) x US-08-630-915A-35 (1-1133)

Qy 91 HisIleGlnLeuAlaLeuAlaLeuArgGluLeuArgSerLeuGluLeuPheArgGlu 110
Db 33 CACAGCAGATGATGGCGGCTTCAAGAGACCAGCAAGAGCTGAGACGGCTTCGGAAG 92
Qy 111 ArgGlnGluGlnArgLysLysTyrGluAlaIleMetAspArgValGlnLysSerLys 130
Db 93 GCACAGAACCCCTGGCCCAAGA----- 116
Qy 131 LeuSerLeuTyrLysLysThrMetGluSerLysLysAlaTyrAspGlnLysCysArgAsp 150
Db 117 -----CTAAGAGGTAGACAGCAGCAAGAAAGCCCATGCACGCTCAAGAG 167
Qy 151 AlaAspAspAla-----GluGlnAlaPheGluArgValSerAlaAsnGlyHisGlnLys 168
Db 168 GAGAGCTGGCTATCTCAGCAGAGGCCAAGCAGCAGACCCATCCCTCAACCTCGAA 227
Qy 169 GlnValGluLysSerGlnAsnLysAlaLysGlnCysLysGluSerAlaThrGluAlaGlu 188
Db 228 CAGCTCAGAAATTCAGACAAATAAGAACTGCAACAGCATGTTCTTAAGACCAA 287
Qy 189 ArgValTyrArgGlnAsnIleGluGlnLeuGluArgValArgThrGluTyrGluGlnGlu 208
Db 288 GAGAGTATGAGAAAGTCCCTGAAGAACTGCACACAGGCAACCCCATGATGAGAAC 347
Qy 209 HisArgThrCysGluAlaPheGlnLeuGlnLysPheAspArgLeuThrIleLeuArg 228
Db 348 ATGAGCAGGTGTTTGAGCAGTGCAGCTTCGAGAGAAACGCTTCCTTCCTCCG 407
Qy 229 AsnAlaLeuTyrValHisCysAsnGlnLeuSerMetGlnCysValLysAsp---AspGlu 247
Db 408 GAGGTTCTCTGAGAGTTTCAAGACCACTAGACCTTCCAATGTGCTGTTACAAGCC 467
Qy 248 LeuTyrGluGluValArgLeuThrLeuGluGlycysAspValGluGlyAspIle 267
Db 468 ATTTCACATGACCTGAGCAGACATCAGACAGCTGATGATGAGAGACCTGAGAGTGG 527
Qy 267 yPheIleGlnSerLysSerThrGlyArg----- 276
Db 528 T-----TTCGAGCAATCAGCGGCGGCATGCGCATGACACTGCGCGAGTTTGAGAG 581
Qy 277 -----GluProProAlaProValProTyrGlnAsnTyrTyrAspArg 290
Db 582 TGTCCCGCAGACCTGATGAGAACCTCAGCCGA-----GAGAG 620
Qy 290 gGluValThrProLeuIleGlySerPro----- 299
Db 621 AAGAGAAAGTCCATGACGCGCTCACCTGACGGGATCAGCAGACGACGACGAGTCT 680

```

OY 300 ----SerIleGlnProSerCysGlyValIleLeuSarphSer-----GlyLeuLeuHi 316
DB 681 CTGCCGAGTAAGCCAGCAGCAGCCCTTATGTCCGAGCAACCCGCCCACTTCGCGAG 740
OY 316 sGlySerProLysThrThrProSer-----AlaProAlaLase 329
DB 741 TCACAGTCCAGCTACACCCCTTCGAGATGAGCAGCAGCAGCGGAGCAGCGTCTAG 800
OY 329 rThrcLthrLeuThrPro-----ThrProG1 338
DB 801 AAGGACGACACTAAGCCAAAATGTGAGCAGCTACGAGAACCCAGCTATCCACAC 860
OY 338 uArgAsnGluLeuValTyrAlaSerIleGluValGlnAlaThrGlnGlnLeuAsn 358
DB 861 GACTGTGTGAGCAGTATGAG-TCTAACACACCCCTTCTCTCCAGCATGGGAGCTC 919
OY 358 rSerAlaGlnAsp-----TyrArgAlaLeuTyrAs 368
DB 920 GAATTCATTCGACGACGACGCCACCTCGGGAGCGAAGTGGAGTCCGGGCGCTGTATGA 979
OY 368 PTyThrAlaGlnAsnSerAspGluLeuAspIleSerAlaGlySpIleLeuAlaVal 388
DB 980 CTATGAGGGGCGACGATGATGAGCTGAGCTTCAAGGCTGGGATGAGCTGACCAAGAT 1039
OY 388 eLeuGluGlyGlnAspGlyTyrPThrValGluArg-----AsnGlyGlnArgGlyPh 406
DB 1040 GGAGGAGCAGATGAGCAGGCGCTGTCGCAAGGAGCGCTTGACACGCGGCAAGTTGGCT 1099
OY 406 eValProGlySerTyrLeuGluLysLeu 415
DB 1100 ATACCCGCAAAATTATGTGAGGCGATC 1127

RESULT 8
US-08-630-915A-29
; Sequence 29, Application US/08630915A
; Patent No. 6309820
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: HOFFMAN, No. 6309820h
; APPLICANT: KAY, Brian K.
; APPLICANT: FOWLER, Dana M.
; APPLICANT: MCCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,915A
; FILING DATE: 03-APR-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELE: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 3345 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-630-915A-29

Alignment Scores:
Pred. No.: 5,34e-09 Length: 3345
Score: 181.00 Matches: 99
Percent Similarity: 35.23% Conservative: 56
Best Local Similarity: 22.50% Mismatches: 123
Query Match: 8,41% Indels: 162
DB: 4 Gaps: 21

US-09-068-377C-1 (1-415) x US-08-630-915A-29 (1-3345)

OY 23 GluValLeuLeuGlnArgLeuLeuAspGlyArgLysMetCysLysAspValGluGluLeu 42
DB 1769 GAATTTATTCGAGAGCGACGACGATATTAAGAGAACTTTTACGCGACGACGAGACTT 1828
OY 43 LeuArgLysAlaGlnAlaGlu----- 50
DB 1829 AACCGGAGGAGCAACAGAGCCGATATTCACCTCGCCGCCACACAGGTGTATCCCGACT 1888
OY 51 -----GluArgTyrGlyLysGluLeuValGlnIle----- 60
DB 1889 CATCATCATTTATCATCATATGAGCGCTTTGGG---GACCTCTCATATATCATATGATACG 1945
OY 61 ----AlaArgLysAlaGlyGlyGlnThrGluMetAsnSerLeuArgThrSerPheAspSer 79
DB 1946 GCCAAAGGAAATCTGGG-----TTAGAGATGAGACCTGCTCGACCCAAATTGAC--- 1996
OY 80 LeuLysGlnGlnThrGluAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaLeuArg 99
DB 1997 TTTAAAGCCGACAGACCTGAG-----GACCTCCCTTCGACAGAGCGA 2038
OY 100 GluGluLeuArgSerLeuGluGluPheArgGluArgGlnLysGlnArgLysLysTyr 119
DB 2039 GAGCTTGT-----TACATCTACACAGATGATGACCAAG---AAGTGTAT 2080
OY 120 GluAlaIleMetAspArgValGlnLysSerLysLeuSerLeuTyrLysLysThMet--- 138
DB 2081 GAAGGT-----GACACCATGGCCGGGATGGGAAATCTTCCACGACCTATATATC 2128
OY 139 -----GluSerLysLysAlaTyrAspGlnLysCysArgAspAlaAspAla 154
DB 2129 GAGCTTCTTCTCCAGCTGAGAGAGGCTCAGCCGAAAGTTGGCACCGGTACAGTTTG 2188
OY 155 Glu-----GlnAlaPheGluArgValSerAlaAsnGlnHisGlnLysGlnValGluLys 172
DB 2189 CATATGAGAGAGCATTCGCAAGTTTAACTTATATGA--GATACACAAAGTGAATG 2245
OY 173 SerGlnAsnLysAlaLysGlnCysLysGluSerAlaThrGluAlaGluArgValTyrArg 192
DB 2246 TCTTCCGAAAGGG----- 2260
OY 193 GlnsnIleGluGlnLeuGluArgAlaArgThrGluThrGluGlnGlnHisArgThrThr 212
DB 2260 ----- 2260
OY 213 CysGluAlaPheGlnLeuGlnLupheAspArgLeuThrIleLeuArgAsnAlaLeuThr 232
DB 2261 -----GAGAGGATCCAGCTGCTCGCAGAGGTG----- 2287
OY 233 ValHisCysAsnGlnLeuSerMetGlnCysValLysAspAspGluLeuTyrGluGluVal 252
DB 2288 -----GATCAGAACGTGAT----- 2302
OY 253 ArgLeuThrLeuGluGlyCysAspValGluGlyAspIleAsnGlyPheIleGlnSerLys 272
DB 2303 -----GAAGGAGAGATTCCTGG----- 2320

```



```

Db      958 -----GAGCCAGGCGCTTACCACTCTGTCTTCTGCCAAGACAGAA 10022
Oy      331 GltThrLeuThrProThrProGluArgasnGluLeuValTyrAlaSerIleGluValGln 350
Db      1003 CCTACCATATGAGATACCCCAAGAGCAGACACCCCTC---TATGAGAACCAACACTGTGTA 1059
Oy      351 AlaThrGluGlyAsnLeuAsnSerIleAlaGlnAspTyr----- 363
Db      1060 CAGCACCAGAGGGGCTGCTCCGACACATTTGACAACTCATGACAGCAGCGGCTTGAGT 1119
Oy      364 -----ArgAlaLeuTyrAspTyrThrAlaGlnAsnSerAspGluLeu 377
Db      1120 GCACAGAGGCGTGTGGCGCCGGGCGCTGTGATGACTACACAGCAGCGATGACACCGAGATC 1179
Oy      378 AspIleSerAlaGlyAspIleLeuAlaValIleLeuGluGlyAluAspGlyTyrTrpThr 397
Db      1180 TCCTTTGACCCCTGAGAGACCTAATCATCAGGCAACGAGTGAATGACAGAGCGTGTGGCGCA 1239
Oy      398 ---ValGluArgAsnGlyGlnArgIlePheValProGlySerTyrLeuGlu 413
Db      1240 GCGTATGGCGCTGAGCGGCGCACTTGTGGCATGTTTCTGTGCCAACTACGTGGAG 1290

RESULT 10
US-08-630-915A-21
/ Sequence 21, Application US/08630915A
/ Patent No. 6309820
/ GENERAL INFORMATION:
/ APPLICANT: SPARKS, Andrew B.
/ APPLICANT: HOFFMAN, No. 6309820h
/ APPLICANT: KAY, Brian K.
/ APPLICANT: FOWLES, Dana M.
/ APPLICANT: MCCONNELL, Stephen J.
/ TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
/ TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
/ NUMBER OF SEQUENCES: 227
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Pennie & Edmonds LLP
/ STREET: 1155 Avenue of the Americas
/ CITY: New York
/ STATE: New York
/ COUNTRY: USA
/ ZIP: 10036-2711
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/630,915A
/ FILING DATE: 03-APR-1996
/ CLASSIFICATION: 536
/ ATTORNEY/AGENT INFORMATION:
/ NAME: MISTOCK, S. Leslie
/ REGISTRATION NUMBER: 18,872
/ REFERENCE/DOCKET NUMBER: 1101-174
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212) 790-9090
/ TELEFAX: (212) 869-8864/9741
/ TELEX: 66141 PENNIE
/ INFORMATION FOR SEQ ID NO: 21:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1531 bases
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA
US-08-630-915A-21

Alignment Scores:
Pred. No.: 4 75e-09 Length: 1531
Score: 176.50 Matches: 97

```

Percent Similarity:	36.98%	Conservative:	72
Best Local Similarity:	21.23%	Mismatches:	171
Query Match:	8.20%	Indels:	117
DB:	4	Gaps:	19

US-09-068-377C-1 (1-415) x US-08-630-915A-21 (1-1531)

```

QY      33 ArgLysMetCysLysAspValGluGlnLeuArgGlnArgAlaGlnIleValArg 52
       ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db     236 AAGAACTGAACCCGAGCGAGGAGAAGGTCTCTGCAGAACTGGGAAGCGGAGAACG 295
QY      53 TyrGlyLysGlnLeuValGlnIleAlaArgLysAlaGlyGln---ThrGlu----- 69
       ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db     296 AAGGACGAGAGATTGAGCAGTGTCTCCACAACACTTCATTAAGCAGCTGACAGAGGCTACC 355
QY      70 -----MetLsnSerIleArgThrSerPheAspSerLeuLysGlnGlnThrAsnVal 87
       :||:|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db     356 CGGCTGCAGAACCATCTTCCCACTATCTGCCCTTTCTTTAAAGCATGACGAA----- 409
QY      88 GlySerAlaHisIleGlnLeuAlaLeuAlaLeuArgGluGlnLeuArgSerLeuGlu 107
       :||:|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db     410 -----GCCCTCAAAGACTGAGTGAGTGTCTTCAGAG 442
QY      108 PheArgGluArgGlnLysGlnArgLysLysTyrgLualalle----- 122
       :||:|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db     443 GTGTAGAGCCCGCAGATGGCTGGCAGAGGATGAAGCAAAACAAGATTGCAGAGCAATGAC 502
QY      123 -----MetAspArgValGlnLysSerLysLeuSerLeuLysLysTrpMetGlu 139
       :||:|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db     503 CTACTCTGATGAGTACACACAGAG-----CTGGTAGCACCGCTCTCTG 550
QY      140 SerLysLysalaeThyr-----AspGlnLysCysArgaspalaaspala 154
       :||:|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db     551 ACATGAGACACCTACCTAGGCCAGTTCCTGATATACATCGCGCATTCGCAAGCGGG 610
QY      155 GluGlnAlaPheGlnArgValSerAlaAsnGlnIleGlnLysGlnValGlu----- 171
       :||:|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db     611 CGGAACCTGATGACATGACAGTGGCGCGGACCACTATGATGCTCTTCAACCGCCAA 670
QY      172 ---LysSerGlnAsnLysAlaLysGlnCysLysGlnSerAlaThrGluAlaGluArgVal 190
       :||:|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db     671 AAGAAAGATGAAGCCAAAATTGCCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 730
QY      191 TyrArgGln---AsnIleGlnLeuGlnIleAlaArgGlnIleArgGlnIleGlnIleHis 209
       :||:|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db     731 TTCGAGAGATGAACGTGATCGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 790
QY      210 ArgThrThrCysGlnAlaPheGlnLeu-----GlnGlu 220
       :||:|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db     791 GGTTCCTATGTCAACACGTTCCAGAGCATCGCGGCTGGAGAAAACTTCCATTAAGAG 850
QY      221 PheaSpArgLeuThrIleLeuArgAsnAlaLeuTrpValHis----- 234
       :||:|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db     851 ATGAGTAAGCTCAATACAGAACTCATATGATGTCTCTGCTGACCTTAGACAGACGACGCG 910
QY      235 CysAsnGlnLeuSerMetGlnCysValLysAspAspGlnLeuTyrGlu----- 250
       :||:|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db     911 AGCAACACCTTCCACAGTCAGAGGCCCAACCCAGCTGACAAATGCCCTGAGAAAGGAACAAG 970
QY      251 -----GluValArgLeuThrIleu 256
       :||:|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db     971 AGCCCGTACCTCTCCAGATGGCTCCCTGCTGCTACCCCTGAGATAGAGGTGAACAT 1030
QY      257 GluGlyCysAspValGluGlyAspIleAsnGlyPheIleGlnSerLysSerThrGlyArg 276
       :||:|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db     1031 GAG-----CCAGACCCGCGCAGTGGGCC 1055
QY      277 GluProProAlaProValProTyrGlnAsnTyrTyrAspArgGluValThrProLeu 296
       :||:|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db     1055 TCACCGGGGGTACCATCCCAAGTCCCACATCTCAGCCAGCAGAGAGGCTCCGAGAGTGTG 1144
QY      297 GlySerProSerIleGlnProSerCysGlyValIleLysArgPheSerGlyLeuLeuHis 316
       :||:|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```



```

Db 108 AAGAGCTACCCGCGGAGAGAGAGTTCCTCCAGAGCTGGGAGGAGGAGATGAGACC 167
OY 52 -----ArgTyrGlyLysGluLeuValGlnIleAla 61
Db 168 AAGATGACGAGTTTGACGACGTGCTCCACAATTTCACAAGACGCTGACGAGGAGACC 227
OY 62 ArgLysAlaGlyGlnThrGluMetAsnSerLeuArgThrSerPheAspSerLeuLys 81
Db 228 CGGCTGCGAAGAGATCTCCGACCTGCTCCGCTCAAAAGCCATGACGAGAGCTTC 286
OY 82 GlnGlnThrGluAsnValGlySerAlaHisIleGlnIleuAlaLeuAlaLeuArgGlu 101
Db 287 CAGAACTGAAATGATGCTGACGAGAGTGTATGAGCCGATTGGCC----- 334
OY 102 LeuArgSerLeuGlnGluPheArgGln-----LysGluGlnArgLys 117
Db 335 -----CGGAGGAGATGAGGAGCAAAAGATCGCAGAGAA 367
OY 118 LysTyrGluAlaIle-MetAspArgValGlnLysSerLysLeuSerLeuTyrLysLys 137
Db 368 CAACGACCTGCTGTGATGATGATTACCAACAGAG-----CTGTGAGACCGAGC 415
OY 137 rMetGluSerLysLysAlaTyr-----AspGlnLysCysArgAspAlaAs 152
Db 416 GCTCTGACATGACACAGTACCTGGCCAGTCCCGACATCAAGTCAGCATTTGCCAA 475
OY 152 PAspAlaGlnAlaPheGluArgValSerAlaAsnGlyHisGlnLysGlnValGlu-- 171
Db 476 GCGGGGCGCAAGCTGTGACACTACGACAGTCCGCGACACATGACGAGTCCCTCAAC 535
OY 172 -----LysSerGlnAsnLysAlaLysGlnCysLysGluSerAlaThrGluAlaG 188
Db 536 TGCCAAAAGAAAGATGAGCAACCAATTCGCCAAGGCCGAGAGAGCTCATCAAGCCCA 595
OY 188 uArgValTyrArgGln--AsnIleGluGlnLeuGlnArgAlaArgThrGluTyrGluG 207
Db 596 GAAGGTGTTGAGATGATGATGATGCTGAGAGAGAGTCCGCTGCTGTGAAGAG 655
OY 207 nGluHisArgThrCysGluAlaPheGlnLeu----- 218
Db 656 CCGGCTAGGTTCTTACGTCAACACGCTCCAGAGCATCCGCGGCTCGAGAGAAACTTCCA 715
OY 219 -GlnGluPheAspArgLeuThrIleLeuArgAsnAlaLeuTyrValHisCysAsnGln-- 237
Db 716 CAGAGAGATGAGCAAGCTCAACCAACCACTCAATGATGTCGTGCGGCTCGAGAGACA 775
OY 238 -----LeuSerMetGlnCysValLysAspAspGluLeuTyrG 250
Db 776 ACACGGAGCAACACCTTCACGCTCAGGCGCCAGCCAGAGAAAGTAACTGTTTC 835
OY 250 uGluValArgLeuThrIleGluGlyCysAspValGluGlyAspIleAsnGlyPheIleG 270
Db 836 GCGGCTGGCGAGAAAGAGACAGACAGGACCGCTCGCAAAAGGAAC-----AA 886
OY 270 nSerLysSerThrArgGluProProAlaProValProTyrGlnAsnTyrTyrAsp 290
Db 887 GAGCCCTTCGCTCCAGATGCTCCCTCCGACCC-----GAGATCAGAGTCAACCA 943
OY 290 gGluValThrProLeuIleGly-----SerProSer-- 300
Db 944 CGAGCCAGAGCCGCGGCGGAGGCCAGCCGCGGAGCCACTTCCCAAGTCCCAATCTCA 1003
OY 301 -----IleGlnProSerCysGlyValIleLysAr 310
Db 1004 GCCAGCAGAGCCCTCGAGGTGCGGAGTGGAGCCCAACCTGCGCGCAGAGCC-----CA 1057
OY 310 gPheSerGlyLeuLeuHisGlySerProLysThrThrProSerAlaProAlaLaserth 330
Db 1058 GAGAGCCAGGAGACTTTCGCAAGTGAAGCAGCCTCAGTCTCTCTCTCTCTCTGTGT 1117
OY 330 rGluThrLeuThrProThrProGluArgAsnGluLeuValTyrAlaSerIleGluValG 350
Db 1118 GGAGACCTTCCAGCAACT-----GTGATGCAACCTGTGAAGGCGG 1159

```

```

OY 350 nAlaThrGlnGlyAsnLeuAsn-----SerSerAlaGlnAspTyrArgAlaLeuTyr 367
Db 1160 CAGTGGGCGCGCGGCTGTGAGCTGCCCCAGGTTTCATGCTTCAAGGTACAGGCCAGCA 1219
OY 367 rAspTyrThrAlaGlnAsnSerAspGluLeuAspIleSerAlaGlyAspIleLeuAla 387
Db 1220 CGACTACACGGCCACTGACACAGACAGAGCTGCAGCTCAAGGCTGTGATGTGTGTGT 1279
OY 387 lIle-----LeuGluGlyGluAspGlyTyr----- 395
Db 1280 GATCCCTTCCAGAACCTTGAGAGCAGAGATGAGAGCTGTGCTCATGGCGTGAAGAGAG 1339
OY 396 -----TyrThrValGluArgAsnGlyGln-----ArgGlyPheValProGlySerTyrLe 412
Db 1340 CGACTGGAGCAGACACAGAAAGCTGGAGAGAGTGCCTTCCCGGAGAACTTCAC 1399
OY 412 uGluLysLeu 415
Db 1400 TGAGAGGCTC 1409

RESULT 15
US-08-919-145-5
: Sequence 5, Application US/08919145
: Patent No. 5958753
: GENERAL INFORMATION:
: APPLICANT: Prendergast, George C.
: TITLE OF INVENTION: Bau, A Binl Interacting Protein, and
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESS: Howson and Howson
: STREET: Spring House Corporate Cntr, P.O. Box 457
: CITY: Spring House
: STATE: Pennsylvania
: COUNTRY: USA
: ZIP: 19477
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/919,145
: FILING DATE: 514
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/025,482
: FILING DATE: 29-AUG-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Kodroff, Cathy A.
: REGISTRATION NUMBER: 33,980
: REFERENCE/DOCKET NUMBER: WST73AUSA
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 215-540-9200
: TELEFAX: 215-540-5818
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1925 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: unknown
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 60..1412
: FEATURE:
: NAME/KEY: mat_peptide
: LOCATION: 60
: US-08-919-145-5
Alignment Scores: 3.68e-06 Length: 1925
Pred. No.:

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: August 3, 2003, 13:20:42 ; Search time 305 seconds
(without alignments)
2807.034 Million cell updates/sec

Title: US-09-068-377c-1
Perfect score: 2152
Sequence: 1 MAAQLOFRDAMFCRDFTAH.....WTVERNGRGFRVGSYLEKL 415

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1439767 seqs, 1031500376 residues
Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n model -DEV=xlh
-Q=/cgn2_1/USFTO.spool/US09068377/rnat_28072003_091319_15426/app_query.fasta.1.583
-DB=published.Applications.NA -QFT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPLC=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.csl -LIST=45 -DOCLIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=1 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09068377 @cgn.1.1.221 @rnat_28072003_091319_15426
-NCPU=6 -ICPU=3 -NO_MAP -LARGOQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DEV.TIMEOUT=120 -WARR.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq1.*
11: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq3.*
13: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1913.5	88.9	1428	15	US-10-067-076-18 Sequence 18, Appl1

2	1910.5	88.8	1428	15	US-10-067-076-19	Sequence 19, Appl1
3	1909.5	88.7	1428	15	US-10-067-076-21	Sequence 21, Appl1
4	1904.5	88.5	1658	15	US-10-067-076-28	Sequence 28, Appl1
5	736	34.2	1606	15	US-10-037-270-926	Sequence 926, App
6	531.5	24.7	914	11	US-09-764-868-408	Sequence 408, App
7	525.5	24.4	473	12	US-09-918-995-31326	Sequence 31326, A
8	390.5	18.1	490	11	US-09-783-590-5577	Sequence 5577, Ap
9	378.5	17.6	488	11	US-09-867-701-10401	Sequence 10401, A
10	310.5	14.4	418	12	US-09-918-995-16410	Sequence 16410, A
11	301	14.0	3255	10	US-09-922-217-683	Sequence 683, App
12	301	14.0	3255	11	US-09-833-263-683	Sequence 683, App
13	301	14.0	3255	15	US-10-025-380-683	Sequence 683, App
14	283.5	13.2	2993	10	US-09-922-217-684	Sequence 684, App
15	283.5	13.2	2993	11	US-09-833-263-684	Sequence 684, App
16	283.5	13.2	2993	15	US-10-025-380-684	Sequence 684, App
17	280	12.3	4091	10	US-09-867-550-1861	Sequence 1861, Ap
18	265.5	12.3	4091	10	US-09-879-957-33	Sequence 33, Appl
19	220.5	10.2	1133	10	US-09-879-957-35	Sequence 35, Appl
20	199.5	9.3	1908	10	US-09-822-849A-102	Sequence 102, App
21	182	8.5	2342	12	US-09-977-418-3	Sequence 3, Appl1
22	182	8.5	2342	12	US-09-977-957-29	Sequence 29, Appl
23	181	8.4	3345	10	US-09-879-957-17	Sequence 17, Appl
24	179.5	8.3	1302	10	US-09-879-957-21	Sequence 21, Appl
25	176.5	8.2	1531	10	US-10-084-817-251	Sequence 251, App
26	175	8.1	1963	15	US-10-175-523-124	Sequence 124, App
27	175	8.1	5796	15	US-10-128-714-7303	Sequence 7303, Ap
28	166.5	7.7	1782	15	US-10-128-714-7303	Sequence 6303, Ap
29	166.5	7.7	1860	15	US-10-128-714-5303	Sequence 5303, Ap
30	166.5	7.7	3860	15	US-09-879-957-13	Sequence 13, Appl
31	166	7.7	1867	10	US-09-954-531-149	Sequence 149, App
32	166	7.7	3248	15	US-09-954-531-149	Sequence 364, App
33	165	7.7	2206	11	US-10-044-090-101	Sequence 101, App
34	165	7.7	2206	11	US-09-952-432A-20	Sequence 20, Appl
35	165	7.7	5652	15	US-09-952-432A-20	Sequence 18, Appl
36	162.5	7.6	1185	11	US-09-954-456-206	Sequence 206, Ap
37	161.5	7.5	1185	11	US-09-954-456-206	Sequence 93, Appl
38	161	7.5	7787	11	US-09-952-432A-22	Sequence 22, Appl
39	160	7.4	1210	11	US-09-952-432A-22	Sequence 214, App
40	159.5	7.4	1185	11	US-09-917-800A-502	Sequence 502, App
41	159	7.4	7420	11	US-09-917-800A-502	Sequence 199, App
42	158	7.3	716	10	US-09-879-957-199	Sequence 191, App
43	154.5	7.2	549	10	US-09-879-957-199	Sequence 5, Appl1
44	153	7.1	1687	10	US-09-866-108-5	
45	152	7.1	4916	10		

ALIGNMENTS

RESULT 1
US-10-067-076-18 Application US/10067076
Sequence 18, Appl1
Publication No. US20030104404A1
GENERAL INFORMATION:
APPLICANT: Wise, Carol A.
FILE OF INVENTION: Genetic Markers for Autoimmune Disorder
TITLE REFERENCE: TEX871/4-006US/36000
CURRENT APPLICATION NUMBER: US/10/067,076
CURRENT FILING DATE: 2002-02-04
PRIOR APPLICATION NUMBER: 60/287,893
PRIOR FILING DATE: 2001-05-01
PRIOR APPLICATION NUMBER: 09/7710,693
PRIOR FILING DATE: 2000-11-08
NUMBER OF SEQ ID NOS: 28
SOFTWARE: PatentIn version 3.2
SEQ ID NO 18
LENGTH: 1428
TYPE: DNA
ORGANISM: Human Nucleic Acid
US-10-067-076-18

Alignment Scores:
Pred. No.: 4.9e-202 Length: 1428
Score: 1913.50 Matches: 366

Percent Similarity: 92.79%
Best Local Similarity: 87.98%
Query Match: 88.92%
DB: 15
Conservative: 20
Mismatches: 29
Indels: 1
Gaps: 1

US-09-068-377C-1 (1-415) x US-10-067-076-18 (1-1428)

```
QY 1 MetMetAlaGlnLeuGlnPheArgAspAlaPheTrpCysArgAspPheThrAlaHisThr 20
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 ATGATGCCCAAGCTGCAAGATGCTTTTGGTGCAGAGGACTTCACAGCCCAACAG 60
QY 21 GLYTGTGValLeuLeuGlnArgLeuLeuAspGlyArgLysMetCysLysAspValGlu 40
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 GGTACAGAGGTGCTGCTGACGGCTTCTGATGCGAGAAAGATGTCMAAGACATGAG 120
QY 41 GluLeuLeuArgGlnArgAlaGlnAlaGluGlnArgGlyGlyLysGluLeuValGlnIle 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 GAGCTACTGAGGCGAGAGGCCAGCGGAGAGCGGATGAGGAGGAGCTGTGCAATC 180
QY 61 AlaArgLysAlaGlyGlnThrGluMetAsnSerLeuArgThrSerPheAspSerLeu 80
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 GCACGGAAGCGAGTGGCCAGAGCATATCACTCCCTGAGGGCTCTTGACTCTTG 240
QY 81 LysGlnGlnThrGluAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaLeuArgGlu 100
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 AACGACCAATGAGAAATGCGGACGCTCACACATCCAGCTGGCCCTGACCTGCTGAG 300
QY 101 GluLeuLeuSerLeuGlnGluPheArgGluArgGlnLysGlnArgLysLysTyGlu 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 GAGCTGGAGGTCTCGAGAGTTTCGTGAGAGCGAGAGGAGCGAGGAGAAAGTATGAG 360
QY 121 AlaIleMetAspArgValGlnLysSerLysLeuSerLeuTyrlLysThrMetGluSer 140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 GCCGTATGACCGCGGTCCAGAAAGAGCAAGCTGCTGCTACAAAGAGGCCATGATGCC 420
QY 141 LysLysAlaArgAspGlnLysCysArgAspAlaAspAlaGluGlnAlaPheGluArg 160
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 AAGAAAGCATACGAGCGAAGTGGCGGAGCGGACGACGCGGAGCGGCTTCGAGCGC 480
QY 161 ValSerAlaAsnGlnHisGlnLysGlnValGlnLysSerGlnAsnLysAlaLysGlnCys 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 481 ATTACGCCCAAGGCCACCGAGAGCGAGTGGAGAGAGTGAAGAAAGCAAGCGCAAGTGC 540
QY 181 LysGluSerAlaThrGlnAlaGluArgValTyrlArgGlnAsnIleGlnLysGluArg 200
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 541 AAGGACTCGGCCACCGAGCGAGCGGCTATACAGCGAGACANTGGCGAGCTGGAGAG 600
QY 201 AlaArgThrGluTrpGluGlnGlnLysArgThrThrCysGluAlaPheGlnLysGln 220
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 601 GTCGGGCTGAGTGGGAGCGAGGACCGGACCGACCTGTAGGCTTTTCAGCTGCAAGAG 660
QY 221 PheAspArgLeuThrIleLeuArgAsnAlaLeuTrpValHisCysAsnGlnLeuSerMet 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 661 TTTCAGCCGCTGACCATTTCTCCGACAGCCCTGTGGGTGCACAGCAACCGCTCTCATG 720
QY 241 GlnCysValLysAspAspGluLeuTyrlGluGlnValArgLeuThrLeuGlnLysAsp 260
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 721 CAGGTGTCTCAAGATGATGAGCTCTACGAGAGAGTGGCTGAGCGTGGAGAGGTGAGC 780
QY 261 ValGluLysPheLeuAsnGlyPheIleGlnSerLysSerThrGlyArgGluProPheAla 280
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 781 ATGAGAGCCGACATCGACATTTTCATCCAGGCAAGAGCGGAGGAGGAGGAGGAGGAG 840
QY 281 ProValProTrpGlnAsnTyrlTyrlAspArgGluValThrProLeuIleGlySerProSer 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 841 CCGGTGCCCTTACCAAGACTATTAAGATCGGAGGTGACCCGCTGACACACACCCCTGGC 900
QY 301 IleGlnProSerCysGlyValIleLysArgPheSerGlyLeuLeuHisGlySerProLys 320
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 901 ATACAGCGCTCTCGGCGCATGAAGAGGTCTCTGAGATGCTGCGACGGAAGTCCAGAG 960
QY 321 ThrThrPro---SerAlaProAlaAlaSerThrGluThrLeuThrProThrProGluArg 339
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 961 ACCACTCTGTTGGAGCTTCTGCTGCCACAGAGACCTGAGACCCCAACCCCGAGCGG 1020
QY 340 AsnGluLeuValTyrlAlaSerIleGluValGlnAlaThrGlnLysAsnLeuAsnSer 359
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1021 AATGAGGTGTCTACACAGCATCGCAGTCCAGAGATACAGGAAACCCGCGCTACCA 1080
QY 360 AlaGlnAspTyrlArgAlaLeuTyrlAspTyrlThrAlaGlnAsnSerAspGluLeuAspIle 379
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1081 GCCCAGAGATACCGGCGGCTCTACGATTAACAGCGCAGAGAACCAATGATGAGACTG 1140
QY 380 SerAlaGlyAspIleLeuAlaValIleLeuGlnGlyLysAspGlyTrpThrValGlu 399
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1141 TCCGCGGAGACATCTCTGAGAGTATCTCGAAGGAGGAGGATGCTGTGACTGTGAG 1200
QY 400 ArgAsnGlyGlnArgGlyPheValProGlySerTyrlLeuGlnLysLeu 415
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1201 AGGACGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1248

RESULT 2
US-10-067-076-19
; Sequence 19, Application US/10067076
; Publication No. US20030104404A1
; GENERAL INFORMATION:
; APPLICANT: Wise, Carol A.
; TITLE OF INVENTION: Genetic Markers for Autoimmune Disorder
; FILE REFERENCE: TEX871/4-006US/36000
; CURRENT APPLICATION NUMBER: US/10/067,076
; CURRENT FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: 60/287,893
; PRIOR FILING DATE: 2001-05-01
; PRIOR APPLICATION NUMBER: 09/710,693
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 19
; LENGTH: 1428
; TYPE: DNA
; ORGANISM: Human Nucleic Acid
US-10-067-076-19

Alignment Scores:
Pred. No.: 1,05e-201 Length: 1428
Score: 1910.50 Matches: 365
Percent Similarity: 92.79% Conservative: 21
Best Local Similarity: 87.74% Mismatches: 29
Query Match: 88.78% Indels: 1
DB: 15 Gaps: 1

US-09-068-377C-1 (1-415) x US-10-067-076-19 (1-1428)
QY 1 MetMetAlaGlnLeuGlnPheArgAspAlaPheTrpCysArgAspPheThrAlaHisThr 20
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 ATGATGCCCAAGCTGCAAGATGCTTTTGGTGCAGAGGACTTCACAGCCCAACAG 60
QY 21 GLYTGTGValLeuLeuGlnArgLeuLeuAspGlyArgLysMetCysLysAspValGlu 40
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 GGTACAGAGGTGCTGCTGACGGCTTCTGATGCGAGAAAGATGTCMAAGACATGAG 120
QY 41 GluLeuLeuArgGlnArgAlaGlnAlaGluGlnArgGlyGlyLysGluLeuValGlnIle 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 GAGCTACTGAGGCGAGAGGCCAGCGGAGAGCGGATGAGGAGGAGCTGTGCAATC 180
QY 61 AlaArgLysAlaGlyGlnThrGluMetAsnSerLeuArgThrSerPheAspSerLeu 80
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 GCACGGAAGCGAGTGGCCAGAGCATATCACTCCCTGAGGGCTCTTGACTCTTG 240
QY 81 LysGlnGlnThrGluAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaLeuArgGlu 100
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 AACGACCAATGAGAAATGCGGACGCTCACACATCCAGCTGGCCCTGACCTGCTGAG 300
QY 101 GluLeuLeuSerLeuGlnGluPheArgGluArgGlnLysGlnArgLysLysTyGlu 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 GAGCTGGAGGTCTCGAGAGTTTCGTGAGAGCGAGAGGAGCGAGGAGAAAGTATGAG 360
```

Oy 121 AlalleleatpavValGlnLysSerLysLeuSerLeuTyrlLysTrnMetLysr 140
 Db 361 GCCCTCATGGACCGGGGTCCAGAAAGCAAGCTGTCCCTCTACAGAAAGGCCATGGATCC 420
 Oy 141 LysLysAlaTyraSpGlnLysCysArgAspAlaAspAspAlaGlnGlnAlaPheGluArg 160
 Db 421 AAGAAAGACATrACGAGAAAGTCCGGGAGCCGGAGCAGCGGGAGCGAGGCTTGCAGCGC 480
 Oy 161 ValSerAlaasnGlnLysGlnLysGlnValGlnLysSerGlnAsnLysAlaLysGlnCys 180
 Db 481 ATTGAGCCCAACGGCCACCAGAAACAGGTGGAGAAGAGTCAAGAAACCAAGCCAGATGC 540
 Oy 181 LysLysSerAlaTrpGlnAlaGlnArgValTyraArgLysAsnLysGlnLysGlnArg 200
 Db 541 AAGGACTTCGGCCACCGAGGAGCAGACGGGGTATACAGCAGAGACATTGGCGAGCTGGAGAG 600
 Oy 201 AlaArGthrGlnTrpGlnGlnGlnLysArgThrThrCysGlnAlaPheGlnLysGln 220
 Db 601 GTCCGGGCTGAGTGGGAGCAGAGACACCGGACCACCTGTGAGGGCTTTCAGCTCAGAGAG 660
 Oy 221 PheAspArgLeuThrLysLeuArgAspAlaLeuTrpValHisCysAsnGlnLysSerMet 240
 Db 661 TTTCACGGCGCTGACCATTTCTCCGAGACGCCCTGTGGGTGCACACAAACACACTTCATG 720
 Oy 241 GlnCysValLysAspAspGlnLysLysTyrlGlnLysValArgLeuThrLysGlnLysCysAsp 260
 Db 721 CAGGTGTGCAGAGTACGATGAGCTTACCAGGAATGGCGCTGACCGCTGGAAAGCTGACAGC 780
 Oy 261 ValGlnLysAspLysGlnLysPheTrpLeuSerLysSerThrGlnArgGlnProProAla 280
 Db 781 ATACAGCGCGCATrACGAGTTTCATCCAGCGCCAAAGAGCAGCGGCACAGACCCCGCT 840
 Oy 281 ProValProTyrlGlnAsnTrpTyrlAspArgGlnValThrProLeuLysGlnSerProSer 300
 Db 841 CCGGTGCCCTTACACAGACTATTACGATCGGGAGGTCAACCCGCTGACCAGAGCCCTGGCC 900
 Oy 301 LysGlnProSerCysGlnValLysLysArgPheSerGlnLysLeuHisGlnLysSerProLys 320
 Db 901 ATACAGCGCGCTCCGGGATGATTAAGAGGTCTCTGGACGTGCTGACAGAGAGTCCAGAG 960
 Oy 321 ThrThrPro--SerAlaProAlaAlaSerThrGlnThrLeuThrProThrProGluArg 339
 Db 961 ACCACTTCGTTGGAGCTTCGTGCGCGCCACAGAGACCCCGACCCCGAGACCGG 1020
 Oy 340 AsnGlnLeuValTyrlAlaSerLysGlnValGlnAlaThrGlnGlnLysAsnLysSerSer 359
 Db 1021 AATAGAGGTGTCTTACACAGCCATGCGACGTGACAGAGATACAGGGAAACCCGGCTTCA 1080
 Oy 360 AlaGlnAspTyrlArgAlaAlaLeuTyrlAspTyrlThrAlaGlnAsnSerAspGlnLysAspLys 379
 Db 1081 GCCCAGAGGATrACCGGGCGCTCTACGATTATrACAGCGCAGAAACCAAGATGAGACTGGAC 1140
 Oy 380 SerAlaGlnAspLysLysLeuAlaValLysLysGlnLysGlnLysGlnLysGlnLysGln 399
 Db 1141 TCCCGCGGAGACATrCCGTGGAGGTATCTCTGAAAGGGAGGTGGCTGGTGGACTGTGGAG 1200
 Oy 400 ArgAsnGlnGlnArgGlnPheValProGlnSerTyrlLeuGlnLysLysLys 415
 Db 1201 AGGAAAGGGGACGCTGGCTTCTGCTCCCTGCTTCTTAACTCGAGAAAGCTT 1248
 RESULT 3
 US-10-067-076-21
 : Sequence 21, Application US/10067076
 : Publication No. US20030104404A1
 : GENERAL INFORMATION:
 : APPLICANT: Wise, Carol A.
 : TITLE OF INVENTION: Genetic Markers for Autoimmune Disorder
 : FILE REFERENCE: TEX871/4-0060S/36000
 : CURRENT APPLICATION NUMBER: US/10/067, 076
 : CURRENT FILING DATE: 2002-02-04
 : PRIOR APPLICATION NUMBER: 60/287,893
 : PRIOR FILING DATE: 2001-05-01

	Prior Application Number: 09/710,693
	Prior Filing Date: 2000-11-08
	Number of SEQ ID NOS: 28
	Software: PatentIn version 3.2
	SEQ ID NO 21
	Length: 1428
	Type: DNA
	Organism: Human Nucleic Acid
	US-10-067-076-21
	Alignment Scores:
	Pred. No.: 1,36e-201
	Score: 1909.50
	Percent Similarity: 92.55%
	Best Local Similarity: 87.74%
	Query Match: 88.73%
	DB: 15
	Gaps: 1
	US-09-068-377C-1 (1-415) x US-10-067-076-21 (1-1428)
QY	1 MetMetLaGlnLeuGlnPheArgAspAlaPheTrpCysArgAspPheThrAlaHisThr 20
Db	1 ATGATGACCCCAAGCTCAGTTCAAAGATGCTTTGGTGTGAGGAGACTTACAGCCACAG 60
QY	21 GtYtYrGluValLeuLeuGlnArgLeuLeuAspGlyArgLysMetCysLysAspValGlu 40
Db	61 GGCTACGAGGTGCTCTCTGAGCGGCTTGTGATGGAGGAAATGTGCAAAGACATAGGAG 120
QY	41 GluLeuLeuArgGlnArgGlnGlnAlaGlnAlaGluGlnArgTyrGlyLysGluLeuValGlnIle 60
Db	121 GAGCTACTGAGGCAAGAGGCGCCAGGCGGAGAGCGGTACGGGAGAGAGCTGTGTCAGATC 180
QY	61 AlaArgLysAlaGlyGlyGlnThrGluMetAsnSerLeuArgThrSerPheAspSerLeu 80
Db	181 GCACGGAAGGACGAGTGGCCAGACGAGATCACTCCTGAGGCGCTCTTGTGACTCCTTGS 240
QY	81 LysGlnGlnThrGluAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaLeuArgGlu 100
Db	241 AAGCGCAAAATCGAATGTGGCGACCTCACATCCACTGACCTGACCCCTGACCTGAG 300
QY	101 GluLeuArgSerLeuGlnGluPheArgGlnArgGlnIleGlnArgLysLysTyrGlu 120
Db	301 GAGCTGCGGAGTCTGAGAGATTCTCTGAGAGGACAGAGAGAGAGAGAGAGATTTGAG 360
QY	121 AlaIleMetAspArgValGlnLysSerLysLeuSerLeuTyrLysLysThrMetGlnSer 140
Db	361 GCCGCTCATGACCGGGTCCAGAAAGACAGATGTGCTCTACAAAGAGCCATGAGATGC 420
QY	141 LysLysLysAlaTyrAspGlnLysCysArgAspAlaAspAspAlaGlnAlaPheGluArg 160
Db	421 AAGAAAGACATACGACAGAGTCCGGGACGGGAGAGACGCGGAGAGGCGCTTCGAGCG 480
QY	161 ValSerAlaAsnGlnHisGlnLysGlnValGlnLysSerGlnAsnLysAlaLysGlnCys 180
Db	481 ATTACCGCCAAACGGCCACCCAGAGGAGGTGGAAACAGCTACAGACAAAGCCAGCATGC 540
QY	181 LysGlnSerAlaThrGlnAlaGlnLysValTyrArgGlnAsnIleGlnLeuGlnGluArg 200
Db	541 AAGAGACTGGCCACCGAGGACAGCGCGGTATACAGGACAGACGACTTCCGACGTGGAGAG 600
QY	201 AlaArgThrGluTrpGluGlnGlnHisArgThrThrCysGluAlaPheGlnLeuGlnGlu 220
Db	601 GTCCGCGGTGAGTGGAGGACGAGGACCGGACCACTGTGAGGCCCTTTCAGCTGCAGAG 660
QY	221 PheAspArgLeuThrIleLeuArgAsnAlaLeuTrpValHisCysAsnGlnLeuSerMet 240
Db	661 TTGAGCCGCGTCACTTCGCAACACCTCTGTGGTGCACACGACCACTCTCATG 720
QY	241 GlnCysValLysAspAspGlnLeuTyrGlnGlnValArgLeuThrLeuGlnGlnCysAsp 260
Db	721 CAGTGTGTCAAGAGATCATGACTCTACAGAGAAAGTGGGCTAGCGTTGGAAGGCTCAGC 780
QY	261 ValGlnGlyAspIleAsnGlyPheIleGlnSerLysSerThrGlyArgGluProAla 280

Db 781 ATAGACCCCGACATCGACAGTTTCATCCAGGCCAAGAGCAGGCGACAGAGCCCGCGCT 840
QY 281 ProvalProtyrGlnAsnTyrrTyraAspArgLysValThrProLeuIleGlySerProser 300
Db 841 CCGGTGCGCTTACCAAGACTATTACGATCGGAGGTCCCGCGTACAGCCCGTGGC 900
QY 301 IleGlnProSerCysGlyValIleLysArgPheSerGlyLeuLeuHisGlySerProLys 320
Db 901 ATACAGCCGTCCTCGGCGCATGATTAAGAGGTCTCTGACTGCTGCACAGGAGTCCCAAG 960
QY 321 ThrThrPro---SerAlaProAlaAlaSerThrGlnThrLeuThrProThrProGluArg 339
Db 961 ACACACTTCCTTGGACGCTTCTGCTGCTCCACAGAGACCTGACCCCGCCGAGCGG 1020
QY 340 AsnGlnLeuValTyrrAlaSerIleGlnValGlnAlaThrGlnGlnIleAsnLeuAsnSer 359
Db 1021 AATGAGGCTCTTACACAGCATCGCAGTGCAGAGATACAGGAAACCCGCGCTCACCA 1080
QY 360 AlaGlnAspTyrrArgAlaLeuTyrrAspTyrrThrAlaGlnAsnSerAspGluLeuAspIle 379
Db 1081 GCCGAGAGATACCGGGGCTCTACGATTATACAGCCGAGAACCATGATGAGCTGGACCTG 1140
QY 380 SerAlaGlyAspIleLeuAlaValIleLeuGlnGlyLysAspGlyTrpThrValGlu 399
Db 1141 TCCGCGGAGACATCTCGAGGTGATCTGGAAGGGAGATGGCTGTGCTGACTGTGGAG 1200
QY 400 ArgAsnGlyGlnArgGlyPheValProGlySerTyrrLeuGlnLysLeu 415
Db 1201 AGGAAGCGGACGCGTCTTCCCTGCTTCTACTGAGAAAGCTT 1248

RESULT 4
US-10-067-076-28
Sequence 28, Application US/10067076
Publication No. US20030104404A1
GENERAL INFORMATION:
APPLICANT: Wise, Carol A.
TITLE OF INVENTION: Genetic Markers for Autoimmune Disorder
FILE REFERENCE: TEX871/4-00605/36000
CURRENT APPLICATION NUMBER: US/10/067,076
CURRENT FILING DATE: 2002-02-04
PRIOR APPLICATION NUMBER: 60/287,893
PRIOR FILING DATE: 2001-05-01
PRIOR APPLICATION NUMBER: 09/710,693
NUMBER OF SEQ ID NOS: 28
SOFTWARE: PatentIn version 3.2
SEQ ID NO 28
LENGTH: 1658
TYPE: DNA
ORGANISM: Human Nucleic Acid
US-10-067-076-28

Alignment Scores:
Pred. No.: 6 03e-201 Length: 1658
Score: 1904.50 Matches: 365
Percent Similarity: 92.55% Conservative: 20
Best Local Similarity: 87.74% Mismatches: 30
Query Match: 88.50% Indels: 1
DB: 15 Gaps: 1

US-09-068-377C-1 (1-415) x US-10-067-076-28 (1-1658)

Db 337 GAGCTACTGAGCGAGAGCGCCAGCGGAGAGCGGTACGAGGAAGAGCTGTGCATATC 396
QY 61 AlArGySaLaGlyGlyGlnThrGlnMetAsnSerLeuArgThrPheAspSerLeu 80
Db 397 GCACGGAAGCAGGTGGCCAGACGAGATACATCCCTGAGGGCGCTCTTGTGACTCTTGC 456
QY 81 LysGlnGlnThrGlnAsnValIleGlySerAlaHisIleGlnIleuAlaLeuAlaArgGlu 100
Db 457 AAGCAGCAATGGAGAAATGTGGCGAGCTGCACATTCAGCTGCGCTGACCTGCGTGAG 516
QY 101 GlnLeuArgSerLeuGlnGlnPheArgGluArgGlnLysGlnGlnArgLysTyrrGlu 120
Db 517 GAGCTGGGAGATCTCCGAGAGTTCTCGAGAGCAGAGAGAGAGAGAGAGAGAGTGTGAG 576
QY 121 AlaIleMetAspArgValGlnLysSerLysLeuSerLeuTyrrLysLysThrMetGluSer 140
Db 577 GCGCTCATGGACCGGCGTCCAGAAAGACCACTCTGCTTACAGAAAGCCAGTGGAGTCC 636
QY 141 LysLysAlaTyrrAspGlnLysCysArgAspAlaAspAspAlaGlnGlnAlaPheGluArg 160
Db 637 AAGAAGCATACAGCAGAAAGTCCCGGAGCGGAGCAGCAGCAGCAGCAGCTTCCAGAGCC 636
QY 161 ValSerAlaAsnGlyHisGlnLysGlnValGlnLysSerGlnAsnLysAlaLysGlnCys 180
Db 697 ATTAGCGCCAAACGCGCACCAAGAGAGGTGGAGAGAGTGAACAACAAAGCCAGCAGTGC 756
QY 181 LysGlnSerAlaThrGlnGlnAlaArgValTyrrArgGlnAsnIleGlnGlnGlnGluArg 200
Db 757 AAGGACTCGGCGCCAGAGGCGAGAGCGGTATACAGCAGAGCAATTCGCGAGCTGGAGAA 816
QY 201 AlaArgThrGlnThrPheGlnGlnGlnIleHisArgThrThrCysGlnAlaPheGlnGlnGlu 220
Db 817 GTCCGGGCTGAGTGGGAGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 876
QY 221 PheAspArgLeuThrIleLeuArgAsnAlaLeuTrpValHisCysAsnGlnLeuSerMet 240
Db 877 TTTCACCGCGCTGACCATTTCTCCCAACGCGCTGTGGTGACAGCAACAGCTCTCCAGT 936
QY 241 GlnCysValLysAspAspGlnLeuTyrrGlnGlnValArgLeuThrLeuGlnGlnCysAsp 260
Db 937 CAGTGTCTCAGAGATATGACCTCTCGAGAGAGTCCGCTGACGCTGGAGAGCTGCAC 996
QY 261 ValGlnGlyAspIleAsnGlyPheIleGlnSerLysSerThrGlyArgGlnProProAla 280
Db 997 ATAGACCGCGACATCGACAGTTTCATCCAGGCAAGACAGCAGGCGACAGCCCGCGCT 1056
QY 281 ProvalProtyrGlnAsnTyrrTyraAspArgLysValThrProLeuIleGlySerProser 300
Db 1057 CCGGTGCGCTTACCAAGACTATTACGATCGGAGGTCCACCCGCTGACAGCAGCCCTGCG 1116
QY 301 IleGlnProSerCysGlyValIleLysArgPheSerGlyLeuLeuHisGlySerProLys 320
Db 1117 ATACAGCCGTCCTCGGCGCATGATTAAGAGCTTCTGAGCTGCTGACAGGAAAGTCCCAAG 1176
QY 321 ThrThrPro---SerAlaProAlaAlaSerThrGlnThrLeuThrProThrProGluArg 339
Db 1177 ACCACTTCGTTGGACACTTCTGCTGCTGCTGCACAGAGACCTGACCCCGCGAGCGG 1236
QY 340 AsnGlnLeuValTyrrAlaSerIleGlnValGlnAlaThrGlnGlyAsnLeuAsnSerSer 359
Db 1237 AATGAGGCTCTTACACAGCATTCGAGTGCAGAGATACAGGAAACCCGCGCTTCACCA 1296
QY 360 AlaGlnAspTyrrArgAlaLeuTyrrAspTyrrThrAlaGlnAsnSerAspGluLeuAspIle 379
Db 1297 GCCAGGATACCGGGGCTCTACGATTATACAGCGCAGAAACCATGATGAGTGCAGCTG 1356
QY 380 SerAlaGlyAspIleLeuAlaValIleLeuGlnGlyLysAspGlyTrpThrValGlu 399
Db 1357 TCCGCGGAGACATCTCGAGAGTGTATCTCGAAGGGAGAGATGGCTGTGAGCTGTGAG 1416
QY 400 ArgAsnGlyGlnArgGlyPheValProGlySerTyrrLeuGlnLysLeu 415
Db 1417 AGGAAGCGGACGCGTCTTCCCTGCTTCTACTGAGAAAGCTT 1464


```
RESULT 8
US-09-783-590-5577
: Sequence 5577, Application US/09783590
: Patent No. US20020110850A1
: GENERAL INFORMATION:
: APPLICANT: Dillon, Patrick J.
: APPLICANT: Haseltine, William A.
: APPLICANT: Li, Haodong
: APPLICANT: Rosen, Craig A.
: APPLICANT: Ruben, Steven M.
: TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
: FILE REFERENCE: PO-16.2C1
: CURRENT APPLICATION NUMBER: US/09/783,590
: PRIOR APPLICATION NUMBER: 08/420,856
: PRIOR FILING DATE: 1995-04-12
: PRIOR APPLICATION NUMBER: 08/346,731
: PRIOR FILING DATE: 1994-11-21
: NUMBER OF SEQ ID NOS: 12485
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 5577
: LENGTH: 490
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc feature
: LOCATION: (116)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (211)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (293)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (295)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (316)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (317)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (326)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (328)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (329)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (331)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (338)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (357)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (359)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (366)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (367)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (371)
```

```
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (375)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (389)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (394)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (395)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (403)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (409)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (415)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (418)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (419)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (424)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (431)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (443)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (468)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (471)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (478)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (479)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (487)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (489)
: OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-5577

Alignment Scores:
Pred. No.: 4.8e-34 Length: 490
Score: 390.50 Matches: 88
Percent Similarity: 79.03% Conservative: 10
Best local Similarity: 70.97% Mismatches: 22
Query Match: 18.15% Indels: 4
DB: 11 Gaps: 1

US-09-068-377C-1 (1-415) x US-09-783-590-5577 (1-490)
Oy 1 MetMetLaGlnLeuGlnPheArgAspAlaPheTrpGysArgAspRphThAlaHisThr 20
Db 21 ATGATGCCCCACGTCAGTTCAGAGATGCCCTTTGGTGCAGGACTTCACAGCCACAGC 80
Oy 21 GATYrGlnuValLeuLeuGlnArgLeuLeuAspLysArgLysMetCysLysAspValGlu 40
:::|||||
```

Db 81 GCTACGAGGTCTGCTGACGCGCTTTCGATGAGNAGAGATGTGCAAGACATGAG 140
QY 41 GtuleuLeuArgGlnArgAlaGlnAlaGlnGluArgTyrGlyLysGlnLeuValGlnIle 60
Db 141 GAGCTACTAAGAGGAGGCGCCGCGGAGAGCGGTAGGAAAGAGAGCTGTGCGCATC 200
QY 61 AlArGlyAlaGlyLysGlnThrGlnMetAsnSerLeuArgThrSerPheAsp-Serie 80
Db 201 GCACGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 260
QY 80 uTysGlnGlnThr-GluAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaArg 100
Db 261 GAAAGCAGCAATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 320
QY 100 LngLLeuLeuArgSerLeu-GluGluPhe---ArgGluArgGlnLysGlnArgLysLys 118
Db 321 GGTAGTNNNGNATTTTNGAGAGGTTTGTGAAGAAANGNAAAGNNGCANGNAGAG 380
QY 119 TyrGln 120
Db 381 TTTGAG 386

RESULT 9

US-09-867-701-10401/C
Sequence 10401, Application US/09867701
Patent No. US20020132237A1
GENERAL INFORMATION:
APPLICANT: Aglate, Paul A.
APPLICANT: Jones, Robert
APPLICANT: Harlocker, Susan L.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
FILE REFERENCE: 210121.497
CURRENT APPLICATION NUMBER: US/09/867,701
CURRENT FILING DATE: 2001-05-29
NUMBER OF SEQ ID NOS: 10912
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 10401
LENGTH: 488
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(488)
OTHER INFORMATION: n - A,T,C or G
US-09-867-701-10401

Alignment Scores:

Pred. No.: 1,02e-32 Length: 488
Score: 378.50 Matches: 81
Percent Similarity: 83.33% Conservative: 4
Best Local Similarity: 79.41% Mismatches: 16
Query Match: 17.59% Indels: 2
DB: 11 Gaps: 1

US-09-068-377C-1 (1-415) x US-09-867-701-10401 (1-488)

QY 315 LeuHtsgLysSerProLysThrPro---SerAlaProAlaAlaSerThrGlnThrLeu 333
Db 488 CTGCGGAGGTCCCAAGACCACTTCGTTGGACGCTTTGTMGGCTCCACAGAGCCCTG 429
QY 334 ThrProThrProGlnArgAsnGlnLeuValTyrAlaSerIleGlnValGlnAlaThrGln 353
Db 428 AC-CCACCCCGGAGCGGAATGAGGTCTTACACAGCCATCGAGTCCAGAGATACAG 370
QY 354 GlysAsnLeuAsnSerSerAlaGlnAspTyrArgAlaLeuTyrAspTyrThrAlaGlnAsn 373
Db 369 GGAACCCGCGCTTACACAGCCAGAGTACCGGGCGCTTACGATTATACAGCCAGAC 310
QY 374 SerAspGlnLeuAspIleSerAlaGlyAspIleLeuAlaValIleLeuGlnGlyLys 393
Db 309 CCAAGATGAGCTGAGACCTGTCCGCGGAGACATCTCTGAGAGTGAATCTGAGAGGAGAT 250

QY 394 GlyTPrPThrValGlnArgAsnGlyGlnArgGlyPheValProGlySerTyrLeuGln 413
Db 249 GGCTGTGACGTGTGAG 190
QY 414 LysLeu 415
Db 189 AAGCTT 184

RESULT 10

US-09-918-995-16410
Sequence 16410, Application US/09918995
Publication No. US20030073623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 16410
LENGTH: 418
TYPE: DNA
ORGANISM: Homo sapiens
US-09-918-995-16410

Alignment Scores:

Pred. No.: 2.76e-25 Length: 418
Score: 310.50 Matches: 61
Percent Similarity: 68.55% Conservative: 24
Best Local Similarity: 49.19% Mismatches: 38
Query Match: 14.43% Indels: 1
DB: 12 Gaps: 1

US-09-068-377C-1 (1-415) x US-09-918-995-16410 (1-418)

QY 2 MetaGlnLeuGlnPheArgAspAlaPheTrpCysArgAspPheThrAlaHisThrGly 21
Db 44 ATGACGCGCTCAGCTGTTAAAGGAACCTTTGGAGTGCAGACATCTCCACCATCGGC 103
QY 22 TyrGlnValLeuLeuGlnArgLeuLeuAspGlyArgLysMetCysLysAspValGlnGlu 41
Db 104 TATGACACATTAATCCACATCTGACACATGCGCGCAAGACTCGCAAGACTTTGAAGAC 163
QY 42 LeuLeuArgGlnArgAlaGlnAlaGlnGluArgTyrGlyLysGlnLeuValGlnIleAla 61
Db 164 TTTCTAAAGAAAGGCGAGCAATTGAGAGGTATGCGCAAGATCTGCTCAACCTCTCT 223
QY 62 ArgLysAlaGly---GlyGlnThrGlnMetAsnSerLeuArgThrSerPheAspSerLeu 80
Db 224 AGGAAGAAAGCCGTGTGACAGCTGGAATTCACACCCCTGAGAGCGGCGCTTGAAGTCTTC 283
QY 81 LysGlnGlnThrGlnAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaLeuArgGln 100
Db 284 AAGCAGCAAGTACATATGTGCGACATGTGCACATTCAGCTTGACAGAGATTAAAGAA 343
QY 101 GtuleuArgSerLeuGlnPheArgGlnArgGlnLysGlnLysGlnArgLysLysTyrGln 120
Db 344 GAGCCACGAGAGGAGAAATTCAGGAGAAAGCAAAACTACAAAGAAAAAGACAGAG 403
QY 121 AlaIleMetAsp 124
Db 404 CTCATTAATGCAAT 415

RESULT 11
US-09-922-217-683
Sequence 683, Application US/09922217
Patent No. US20020076414A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun

```

: APPLICANT: Lodes, Michael J.
: APPLICANT: Secrist, Heather
: APPLICANT: Benson, Darin R.
: APPLICANT: Meagher, Madeleine Joy
: APPLICANT: Stolk, John A.
: APPLICANT: Wang, Tonglong
: APPLICANT: Jiang, Yujia
: APPLICANT: Smith, Carole Lynn
: APPLICANT: King, Gordon E.
: APPLICANT: Wang, Aijun
: APPLICANT: Clapper, Jonathan D.
: TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
: FILE REFERENCE: 210121.471C13
: CURRENT APPLICATION NUMBER: US/09/922.217
: NUMBER OF SEQ ID NOS: 1124
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO: 683
: LENGTH: 3255
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-922-217-683

Alignment Scores:
Pred. No.: 6e-23 Length: 3255
Score: 301.00 Matches: 113
Percent Similarity: 41.39% Conservative: 89
Best Local Similarity: 23.16% Mismatches: 188
Query Match: 10 Indels: 99
Gaps: 17

US-09-068-377c-1 (1-415) x US-09-922-217-683 (1-3255)

QY 9 AspaAlpheTtrPcySarGspHethrAlaHisThGly---TyrGluValLeuLeuGln 27
    |||:|||||
    254 GACAGCTTCG-----GAGCTCGGAACTACAAGCCGACTGTGAAG 295
    |||:|||||

QY 28 ArgLeuLeuAspGlyArgLysMetCysLysAspValGluLeuLeuLeuArgGlnAla 47
    |||:|||||
    296 CGGATGCGAGATGGCCACCCCTGTGCGAGCCCTCATGAACGTGCGATGAGCGGCG 335
    |||:|||||

QY 48 GlnAlaGluGluArgTyrGlyLysGluLeuValGlnLeuAlaArgLys----- 63
    |||:|||||
    356 CGCATGAGAGAGCGGATGCGAGCAGCTCAGTCAAGTGGGCGCGCGCTGAGGAGCTT 415
    |||:|||||

QY 64 -----AlaGlyGlyGlnThrGluMetAsnSerLeuArgThrSerPheAspSerLeu 81
    |||:|||||
    416 GTGGAGAAAGGCGCCAG-----TACGGGAGCCGTGGAGAGCGCTGATGGCTTCATG 469
    |||:|||||

QY 82 GlnGlnThrGluAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaLeuArgGlu 101
    |||:|||||
    470 TCCGAGGCGAGAGGGGTGAGCGAGCTGACCTGAGGTGAAGCCCTCATGATGAACAT 529
    |||:|||||

QY 102 LeuArgSerLeuGluLupheArgLuarGlnLysGlu----- 114
    |||:|||||
    530 -----GACTTCGAGAAGATCAAGAACTGSCAGAGAAACCTTTCACAGCAGATGATG 583
    |||:|||||

QY 115 -----GlnArgLysLysTyrGluAlaIleMetAspArgValGlnLysSerLys 130
    |||:|||||
    584 GGGCGCTTCAAGAGAGCAAGGAAAGCTTTCGAGAGCGCTTCGAGAGCAGACAGCCCTG 643
    |||:|||||

QY 131 LeuSerLeuTyrLysThrMetGluSerLysAlaLeuTyrGlnLysCysArgAsp 150
    |||:|||||
    644 GCCAAGAGCTGAAAGAGGTAGAGAGCAAGCAAGAGCAACCATCTTCACACCCCTGAA 703
    |||:|||||

QY 151 AlaAspAspAla-----GlnGlnAlaPheGluArgValSerAlaAsnGlnLysGln 168
    |||:|||||
    704 GAGAACTGGCTATCTCAGAGAAAGCAAGCAAGCAAGCAACCATCTTCACACCCCTGAA 763
    |||:|||||

QY 169 GlnValGlnLysSerGlnAsnLysAlaLysGlnCysLysGlnSerAlaThrGlnAlaGlu 188
    |||:|||||
    764 CAGCTCAAGAAATTGCAAGACAAATAAGAAAGTGAAGCAAGCAAGATGTTTAAAGACCAA 823
    |||:|||||
```

```

QY 189 ArgValTyrArgGlnAsnIleGlnLeuGluArgAlaArgThrGluTrpGluGlnGlu 208
    |||:|||||
    824 GAGAAATATGAGAGTCCCTCAAGAACTGACAGGCGCACCCAGTACATGAGAGAAC 883
    |||:|||||

QY 209 HisArgThrCysGluAlaPheGlnLeuGlnLupheAspArgLeuThrIleLeuArg 228
    |||:|||||
    884 ATGAGCAGAGGTGTTGAGCAGAGTGCAGCTTGCAGAGAGAAAGCCCTTCCTTCCTCGG 943
    |||:|||||

QY 229 AsnAlaLeuTyrValHisCysAsnGlnLeuSerMetGlnCysValLysAsp---AspGlu 247
    |||:|||||
    944 GAGGTTCGCTGGAGGTTCAAGAACCACTTAACCTGTCATGCTGCTGCTGTTCAAAAGCC 1003
    |||:|||||

QY 248 LeuTyrGlnLuvAlaArgLeuThrLeuGlnLugLysAspValGlnLysPheAsnGly 267
    |||:|||||
    1004 ATTACCATGACCTGGAGCAGAGCATGACAGACAGCTGATGAGGAGCTGAGAGTGG 1063
    |||:|||||

QY 268 PheIleGlnSer-LysSer-----ThrGlyArg----- 276
    |||:|||||
    1064 TTCGAGCCAAATCACGGGCGCAGGCATGACATGACCTGCGCGAGTTGAGAGTGGTCC 1123
    |||:|||||

QY 277 -----GluProProAlaProValProTyrGlnAsnTyrTyrAspArgGluVal 292
    |||:|||||
    1124 GCAAGACCTGATTCGAAACCTCAGCCGGA-----GAGAGAAAGAG 1162
    |||:|||||

QY 292 LThrProLeuIleGlySerPro-----Se 300
    |||:|||||
    1163 AAGCCACTAGACGGCTTACCCCTGACGCGGATCAACCAAGAGCGACAGCTTTTGGCCG 1222
    |||:|||||

QY 300 rIleGlnProSerCysGlyValIleLysArgPheSer-----GlyLeuLeuHisGlySe 318
    |||:|||||
    1223 AGTAAGCCCAAGCAGACACCTTAATGTCCGAGACACCCCGCCAGCTGGCGAGTCAAG 1282
    |||:|||||

QY 318 rProLysThrThrProSer-----AlaProAlaAlaSerThrGln 331
    |||:|||||
    1283 TCCAGCTACAAACCCCTTCGAGATGAGAGAGACAGACAGCGGACACCTCAGTGAAGAAAGAG 1342
    |||:|||||

QY 331 uThrLeuThrPro-----ThrProGluArgAs 340
    |||:|||||
    1343 GACATTAAAGCCCAAAATGTGAGCAGCTACGAGAGACCCAGAGCTTATCCACCGACTGG 1402
    |||:|||||

QY 340 nGluLeuValTyrAlaSerIleGluValGlnAlaThrGlnGlyAsnLeuAsnSerSerAl 360
    |||:|||||
    1403 TCAGACATGAG--TCATACAAACCCCTTCCTCAGAGATGCCAATGGGAGCTGATGCC 1461
    |||:|||||

QY 360 aGlnAsp-----TyrArgAlaLeuTyrAspTyrTh 370
    |||:|||||
    1462 ATTGACGAGCAGAGCCACTTCGGGAGAGAAAGTCGAGTCCGGCCCTGTATGACTATGA 1521
    |||:|||||

QY 370 rAlaGlnAsnSerAspGluLeuAspLieserAlaGlyAspIleLeuAlaValIleLeuGln 390
    |||:|||||
    1522 GGGCAGAGCAGCATGATGCTGAGCTTCAAGGCTGGGATGAGCTGACCAAGATGAGAGA 1581
    |||:|||||

QY 390 uGlyGluAspGlyTyrProThrValGluArg-----AsnGlyGlnArgGlyPheValPr 408
    |||:|||||
    1582 CGAGGATGAGCAGAGGCGTGGTGAAGGAGCCTTGGAAAGGGAATTTGGCTATGCC 1641
    |||:|||||

QY 408 oGlySerTyrLeuGluLysLeu 415
    |||:|||||
    1642 GGCAGAAATATATGTGAGGCGCATC 1663
    |||:|||||

RESULT 12
US-09-833-263-683
: Sequence 683, Application US/09833263
: Patent No. US20020110547A1
: GENERAL INFORMATION:
: APPLICANT: Wang, Aijun
: APPLICANT: Clapper, Jonathan D.
: APPLICANT: Stolk, John A.
: APPLICANT: Meagher, Madeleine J.
: TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
: FILE REFERENCE: 210121.471C12
```



```

; NUMBER OF SEQ ID NOS: 1129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 683
; LENGTH: 3255
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-025-380-683

```

Alignment Scores:

```

Pred. No.: 6e-23      Length: 3255
Score: 301.00      Matches: 113
Percent Similarity: 41.39%      Conservative: 89
Best Local Similarity: 23.16%      Mismatches: 188
Query Match: 13.99%      Indels: 99
DB: 15      Gaps: 17

```

US-09-068-377c-1 (1-415) x US-10-025-380-683 (1-3255)

```

QY 9 AspalaphetPCysArgrsPheThrAlaHisThrGly--TyrGluValLeuLeuGln 27
    |||:|||||
DB 254 GACAGCTTGTG-----GAGGTGGGAACTACAAAGCGGAGCTGTGAG 295

QY 28 ArgLeuLeuAspGlyArgLysMetCysLysAspValGluGluLeuArgGlnArgAla 47
    |||:|||||
DB 296 CGAGTCGACGATGGCCACCCCTGTCTGACGACCTCATGAACTGCCTCATGAGCGGCG 355

QY 48 GlnAlaGluGlnArgGlyArgLysGluLeuValGlnLeuAlaArgLys----- 63
    |||:|||||
DB 356 CGCATCGAAGAGCGGTATCGCAGCAGCTCATGAGTGGCCGGCGGTGAGCGCAGCTT 415

QY 64 -----AlaGlyGlyGlnThrGluMetAsnSerLeuArgThrSerPheAspSerLeuLys 81
    |||:|||||
DB 416 GTGGAGAAAGGGGCCAG-----TACGGGACCGTGAGAGAGCCGTGATGGCTTCATG 469

QY 82 GlnGlnThrGlnAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaLeuArgGlu 101
    |||:|||||
DB 470 TCCGAGGCGAGAGGGGTGAGCGAGCTGCACCTCGAGGTGAGAGCCCTCATGATGAACGAT 529

QY 102 LeuArgSerLeuGluGlnPheArgGluArgGlnLysGlu----- 114
    |||:|||||
DB 530 -----GACTTCGAGAGATCAAGAACTGGCAGAGAGAGCCCTTCACAAAGCAGATGATG 583

QY 115 -----GlnArgLysLysTyrGluAlaIleMetAspArgValGlnLysSerLys 130
    |||:|||||
DB 584 GCGCGCTTCAGAGAGCAAGAGAGCTGAGACGGCTTCGAGAGCAGACAGAGCCCTG 643

QY 131 LeuSerLeuTyrLysLysThrMetGluSerLysLysAlaTyrAspGlnLysCysArgAsp 150
    |||:|||||
DB 644 GCCAAGAGCTGAAAGAGGTAGAGAGCAAGAGAGAGAGCCCATGACGCTGCAAAAGAG 703

QY 151 AlaAspAspAla-----GluGlnAlaPheGlnArgValSerIleAsnGlyHisGlnLys 168
    |||:|||||
DB 704 GAGAAAGCTGGCTATCTCAGCAGAGAGCAAGCAGAGAGCCCATCTTCACCTGGA 763

QY 169 GlnValGluLysSerGlnAsnLysAlaLysGlnCysLysGluSerAlaThrGlnAlaGlu 188
    |||:|||||
DB 764 CAGCTCAAGAAATTCAGAGCAAAATATGAAAGTGCAGAGAGATGTTCTTAAGACCAAA 823

QY 189 ArgValTyrArgGlnAsnIleGluGlnLeuArgAlaArgThrGluTyrGluGlnGlu 208
    |||:|||||
DB 824 GAGAAATAGAGAGAGCTCAAGAGAACTGCAGCAGAGAGAGCCCATGATGAGAGAG 883

QY 209 HisArgThrThrCysGluAlaPheGlnLeuGlnGluPheAspArgLeuThrIleLeuArg 228
    |||:|||||
DB 884 ATGGAGCAGAGTGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 943

QY 229 AsnAlaLeuTyrValHisCysAsnGlnLeuSerMetGlnCysValLysAsp---AspGln 247
    |||:|||||
DB 944 GAGGTTCCTGAGAGGTTCAGAGAGCAGCTAAACCTTCACAGTGGCTGTTCACAAAGC 1003

QY 248 LeuTyrGluGluValArgLeuThrIleGluGlnGlyCysAspValGlnLysPheAsnGly 267
    |||:|||||
DB 1004 ATTACCATGACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1063

```

```

QY 268 PheIleGlnSer-LysSer-----ThrGlyArg----- 276
    |||:|||||
DB 1064 TTCGAGCCAAATACAGCGCCAGAGCATGCCATCAACTGCGCGAGTTTGAAGAGTGTC 1123

QY 277 -----GluProAlaProValProTyrGlnAsnTyrTyrAspArgGluVal 292
    |||:|||||
DB 1124 GCAGACCTGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1162

QY 292 LThrProLeuIleGlySerPro-----Se 300
    |||:|||||
DB 1163 AAGCCCATCGACGCGCTTCACCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1222

QY 300 rIleGlnProSerCysGlyValIleLysArgPheSer-----GlyLeuLeuHisGlySe 318
    |||:|||||
DB 1223 AGTAACCCAGCAGCAGCCCTTAAATGTCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1282

QY 318 rProLysThrThrProSer-----AlaProAlaAlaSerThrGln 331
    |||:|||||
DB 1283 TCCAGCTACAGACCCCTTCGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1342

QY 331 uThrLeuThrPro-----ThrProGluArgAs 340
    |||:|||||
DB 1343 GACATTAAAGCCCAAAATGTGACAGCATACAGAGAGAGAGAGAGAGAGAGAGAGAG 1402

QY 340 nGluLeuValTyrAlaSerIleGluValGlnAlaThrGlnGlyAsnLeuAsnSerSerAl 360
    |||:|||||
DB 1403 TCAGAGCATGAG-TCTAACACCCCTTCCTCAGAGATGCCAATGGGAGCTGATATCC 1461

QY 360 acLnaSp-----TyrArgAlaLeuTyrAspTyrTh 370
    |||:|||||
DB 1462 ATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1521

QY 370 rAlaGlnAsnSerAspGluLeuAspIleSerAlaGlyAspIleLeuAlaValIleLeuGln 390
    |||:|||||
DB 1522 GGGGCGAGAGAGATGATGAGTGTGAGCTTCAAGGCTGGGAGATGAGCAGCAAGATGAGGA 1581

QY 390 uGlyGluAspGlyTyrThrThrValGluArg-----AsnGlyGlnArgGlyPheValPr 408
    |||:|||||
DB 1582 CGAGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1641

QY 408 ogLysSerTyrLeuGlnLysLeu 415
    |||:|||||
DB 1642 GCCAATTAATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1663

RESULT 14
US-09-922-217-684
; Sequence 684, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yuguang
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922.217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 684
; LENGTH: 2993
; TYPE: DNA
; ORGANISM: Mus musculus

```

US-09-922-217-684

Alignment Scores:

Pred. No.:	4.61e-21	Length:	2993
Score:	283.50	Matches:	112
Percent Similarity:	39.29%	Conservative:	88
Best Local Similarity:	22.00%	Mismatches:	169
Query Match:	13.17%	Indels:	141
		Gaps:	16

US-09-068-377C-1 (1-415) x US-09-922-217-684 (1-2993)

```

QY 9 AspaLabhPrpCysArgAspPheThrAlaHisThrGly---TyrGluValLeuLeuGln 27
DB 84 GACAGCTTCTGCG-----CAGCTTGGGAACCTCAACAGCAGCTGTAG 125

QY 28 ArgLeuLeuAspGlyArgGlyMetCysLeysAspValGluGluLeuLeuArgGlnAla 47
DB 126 CGGATTGACGATGGCCACCGCTGTGTGTCACCTCATGAACCTGTCATGACGGGCA 185

QY 48 GlnAlaGluGluArgTyrGlyLeuValGlnLeuAlaArgLys----- 63
DB 186 CGCATTCAGAAAGCGCTATGCACACAGCTCAGTGGGCCACCTGAGGACACTG 245

QY 64 -----AlaGlyGlyGlnThrGluMetAsnSerLeuArgThrSerPheAspSerLeuLys 81
DB 246 GTACAGAAAGGACACAG-----TATGGACCTGGAGAAAGCGCTGATAGTGCATG 299

QY 82 GlnGlnThrGluAsnValGlySerAlaHisLeuLeuAlaLeuAlaLeuArgGlnGlu 101
DB 300 TCTGAACACAGAGAGGAGTGAAGTGAACCTGGAATGAAAGCCATCATGATGAATGAA 359

QY 102 LeuArgSerLeuGluGluPheArgGluArgGlnLysGlu----- 114
DB 360 -----GACTTGAAGATCAACAGAACTGCAGAGAGAGCGCTTTCACAAAGCAGATGATG 413

QY 115 -----GlnArgLysLysTyrGluAlaLeuMetAspArgValGlnLysSerLys 130
DB 414 GAGAGCTTCAAGACCAAAAGAGCAGAGATGGCTTTCGGAAGCGCCGAAACCCCTGG 473

QY 131 LeuSerLeuArgLysLysThrMetGluSerLysLysAlaTyrAspGlnLysCysArgAsp 150
DB 474 GCCAAGAAAGCTGAAGAGGTGAAGCGCAAAAGAGGCGCACACAGCGCTGAAGAGAG 533

QY 151 AlaAspAspAla-----GluGlnAlaPheGluArgValSerAlaAsnGlnHisGlnLys 168
DB 534 GAGAGCTGGCCATCTCCCGGAAGCCACACAGCAGCAGATCCATCCCAACCCCTGAG 593

QY 169 GlnValGluLysSerGlnAsnLysAlaLysGlnCysLysGluSerAlaThrGluAlaGlu 188
DB 594 CAGCTGAAGAAAGCTGCAAGCAAGATGAAAAATGCAAAACAGACGCTTAAAGACCAAG 653

QY 189 ArgValTyrArgGlnAsnLeuGluGlnLeuArgAlaArgThrGluTyrGluGlnGlu 208
DB 654 GACAAGATATGAAAGTCCCTGAAGAGACTGTGATGACCAACACCCCGTACATGGAAGAC 713

QY 209 HisArgThrArgCysGluAlaPheGlnLeuGlnGluPheAspArgLeuThrIleLeuArg 228
DB 714 ATGAGCAGAGCTGTTCAGACAGCTGCACACAGCTTTGAAGAAGAGCCCTTCCTCCGG 773

QY 229 AsnAlaLeuTyrValHisCysAsnGlnLeuSerMet----- 240
DB 774 GAGATTCTCTGAGAGTTCAAGAACACTGTGATCTGCATGATGGCTAGCTATAAAGC 833

QY 241 -----GlnCysValLysAspAspGluLeuTyrGluGluValArg--- 253
DB 834 ATTATCCGGAGAGCTGAGCAGACAGCATCAAGCAGACAGATCGGTAGAGACCTGAGGTGG 893

QY 253 ----- 253
DB 894 TTCGGGCTAACCATGGGCGCAGCATGGCTATGTAAGTGGCCACAGTTTGAGAGTGTCT 953

QY 254 -----LeuThr 255

```

```

DB 954 GCAGATCTGAATGCACTCTCAGCGGAGAGAGAGAAGAGGCTGTTGACGGTGTACCC 1013
QY 256 LeuGluGlyCysAspValGluGlyAspIleAsnGlyPheIleGlnSerLys----- 272
DB 1014 CTAAACAGGATCAACCCAGACAGAGGTCAGCTGTGA-----CAGAACAAGCCCTGGCAGC 1067
QY 273 -----Ser-ThrGlyArgGluProPheAlaProValProTyrGlnAsnTyrTyrAsp 290
DB 1068 AACCTTAGTGTCCCGACCAACCCCGCCAGTCCACGACGATTAC----- 1110
QY 290 GlnValThrProLeuIleGlySerProSerIleGlnProSerCysGlyValIleLysArg 310
DB 1111 -----AGTCCAGCTACAAACCCCTTCG----- 1131
QY 310 gPheSerGlyLeuLeuHisGlySerProLysThrThrProSerAlaProAlaLaserThr 330
DB 1132 -----AGGACGAGGACGACAGGGGACAGCATCATCATGAGAAC 1169
QY 330 rGluThrLeuThrPro-----ThrProGluArg 339
DB 1170 GAGACACTTAAGCCAAATAATGTCAGACGCTATGAGAAGCTCAGACTTACCCACTGAC 1229
QY 339 gAsnGluLeuValTyrAlaSerIleGluValGlnAlaThrGlnLysAsnLeuAsnSerSe 359
DB 1230 TGCTCTGATGATGAG-TCTAACAACCTTCTCTCCACGATGCCAAGCGGATTCGAA 1288
QY 359 rAlaGlnAsp-----TyrArgAlaLeuTyrAsp 369
DB 1289 CCCATTGATGAGACGACACGACCTCAGAGACAGAGAGTGGCGGCTTCATATGAC 1348
QY 369 rThrAlaGlnAsnSerAspGluLeuAspIleSerAlaGlyAspIleLeuAlaValIle 389
DB 1349 TGAGCGGCGACAGACATGATGATGAGCTTCAAGCGCTGGGAGATGACATGCACAGATAGA 1408
QY 389 uGluGlyGluAspGlyTyrThrValGluArg-----AsnGlyGlnArgGlyPheVal 407
DB 1409 GCATGAAGATGACAGAGGTTGTGCAAGGAGCGCTTAAAGACGCGCAGGTTGGCTATA 1468
QY 407 lProGlySerTyrLeuGluLysLeu 415
DB 1469 CCCAGCCCACTATGTGAGGCTATC 1493

RESULT 15
US-09-833-263-684
; Sequence 684, Application US/09833263
; Patient No. US20020110547A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Stolk, John A.
; APPLICANT: Meagher, Madeleine J.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; FILE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121, 471612
; CURRENT APPLICATION NUMBER: US/09/833,263
; NUMBER OF SEQ. ID NOS.: 1093
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 684
; LENGTH: 2993
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-833-263-684

Alignment Scores:
Pred. No.: 4.61e-21 Length: 2993
Score: 283.50 Matches: 112
Percent Similarity: 39.29% Conservative: 88
Best Local Similarity: 22.00% Mismatches: 169
Query Match: 13.17% Indels: 141
DB: 11 Gaps: 16

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 3, 2003, 13:49:47 ; Search time 3433 Seconds

(Without alignments)
3518.108 Million cell updates/sec

Title: US-09-068-377C-1

Perfect score: 415

Sequence: 1 MMAQLQFRDAFCRCDFTAHT.....WTVERNQGRFVPGSYLEKL 415

Scoring table: OLIGO

Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2054640 segs, 14551402878 residues

Word size: 1

Total number of hits satisfying chosen parameters: 4106490

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=rtlh
-O/cgnt2_1/USFTO.spool/US09068377/rnatc_28072003_091402_15914/app_query.fasta.1.583
-DB=GenEmbl -QFMT=fastap -SUFFIX=olip2n.rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEADSIZE=500 -MINKEN=0 -MAXLEN=2000000000
-USFR=US09068377 @CCN 1.1.2496 @rnatc_28072003_091402_15914 -NCFU=6 -ICPU=3
-NO_MMAP -LARGEODRERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -Fgapop=6
-Fgapext=7 -Ygapop=60 -Ygapext=60 -Delop=6 -Delext=7

Database :

1: GenEmbl:*
2: gb_ba:*
3: gb_htg:*
4: gb_in:*
5: gb_com:*
6: gb_ov:*
7: gb_pat:*
8: gb_ph:*
9: gb_pl:*
10: gb_pr:*
11: gb_ro:*
12: gb_sts:*
13: gb_sy:*
14: gb_un:*
15: gb_vi:*
16: em_ba:*
17: em_fun:*
18: em_hum:*
19: em_in:*
20: em_mu:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pl:*
27: em_ro:*
28: em_sts:*
29: em_un:*

29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	415	100.0	1692	10 MM087814	U87814 Mus musculus
2	415	100.0	2100	6 AR108626	AR108626 Sequence
3	45	10.8	44437	2 AC100102	AC100102 Mus muscu
4	45	10.8	72661	2 AC123729	AC123729 Mus muscu
5	45	10.8	161466	2 AC116689	AC116689 Mus muscu
6	45	10.8	173635	2 AC123432	AC123432 Rattus no
7	45	10.8	200703	2 AC123482	AC123482 Rattus no
8	35	8.4	57518	2 AC101133	AC101133 Mus muscu
9	34	8.2	173635	2 AC123432	AC123432 Rattus no
10	34	8.2	200703	2 AC123482	AC123482 Rattus no
11	33	8.0	44437	2 AC100102	AC100102 Mus muscu
12	33	8.0	57518	2 AC101133	AC101133 Mus muscu
13	31	7.5	14382	9 AF038602	AF038602 Homo sapi
14	31	7.5	14382	9 AF038603	AF038603 Homo sapi
15	31	7.5	1656	9 HSU94778	U94778 Human PEST
16	31	7.5	1781	9 BC008602	BC008602 Homo sapi
17	29	7.0	562	9 IROSE015	AI101010 Homo sapi
18	29	7.0	131163	2 AC107883	AC107883 Homo sapi
19	29	7.0	135317	2 AC090181	AC090181 Homo sapi
20	26	6.3	154934	2 AC090177	AC090177 Homo sapi
21	25	6.0	72661	2 AC123729	AC123729 Mus muscu
22	16	3.9	258	11 G11561	G11561 human SFS S
23	10	2.4	14381	3 DROCCHC64C	L23195 Drosophila
24	10	2.4	75967	2 AC014970	AC014970 Drosophila
25	10	2.4	110506	2 AC008374	AC008374 Homo sapi
26	10	2.4	163709	3 AC108877	AC108877 Drosophila
27	10	2.4	183680	9 AC098859	AC098859 Homo sapi
28	10	2.4	187437	3 AC023751	AC023751 Drosophila
29	10	2.4	188697	2 AC125577	AC125577 Rattus no
30	10	2.4	301769	3 AE003482	AE003482 Drosophila
31	9	2.2	497	4 BOVCRRY01	M16895 Bovine gamm
32	9	2.2	619	4 BTRCRY01	X01036 Calif lens m
33	9	2.2	624	4 BOVCRRYGBA	M16894 Bovine gamm
34	9	2.2	1321	6 A02597	A02597 Synthetic (
35	9	2.2	2185	10 RRMETIII	Y08235 Rattus us
36	9	2.2	4530	1 BACBIOXWF	W29291 B.sphaericu
37	9	2.2	34380	2 AP005530	AP005530 Homo sapi
38	9	2.2	38181	2 AP005596	AP005596 Homo sapi
39	9	2.2	59179	2 AC090394	AC090394 Homo sapi
40	9	2.2	62988	2 AL731680	AL731680 Human DNA
41	9	2.2	65147	2 AC101675	AC101675 Mus muscu
42	9	2.2	67309	2 AL157932	AL157932 Human DNA
43	9	2.2	68870	2 AC130703	AC130703 Homo sapi
44	9	2.2	72552	2 AC032032	AC032032 Homo sapi
45	9	2.2	95283	9 HS599F21	AL035662 Human DNA

RESULT 1

ALIGNMENTS

MM087814 1692 bp mRNA linear ROD 30-OCT-2001
LOCUS Mus musculus PEST phosphatase interacting protein mRNA, complete
DEFINITION cds
ACCESSION U087814
VERSION 087814.1 GI:1857711
KEYWORDS
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 1692)
AUTHORS Spencer, S., Dowbenko, D., Cheng, J., Li, W., Brush, J., Utzig, S., Simanis, V., and Lasky, L. A.
TITLE PSTPIP: a tyrosine phosphorylated cleavage furrow-associated protein that is a substrate for a PEST tyrosine phosphatase
JOURNAL J. Cell Biol. 138 (4), 845-860 (1997)
MEDLINE 97412181
PUBMED 9265651
REFERENCE 2 (bases 1 to 1692)
AUTHORS Lasky, L.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-1997) Molecular Oncology, Genentech, 460 Pt. San Bruno Blvd., San Francisco, CA 94080, USA
FEATURES
source
1. 1692
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_type="hematopoietic progenitor cells"
274. 1521
/note="similar to S. pombe Cdc15p, a protein associated with formation of the cleavage furrow during cytokinesis; actin interacting protein with C-terminal SH3 domain; PSTPIP"
/codon_start=1
/product="PEST phosphatase interacting protein"
/protein_id="AAB48483.1"
/db_xref="GI:1857711"
/translation="MMAOLFADFAFCMDPFAHTGYEVLRLDGRKMKCVDELLR ORAOEERYGKELVIAKAGOTEMSLRTPSKMOTGVNSAHQLALAREEL RSLPEERROKRRKYEAIMDRYKSLSTXKTMESKRAYDCKRDADEAFER VSAHGOKOYKESONKAKOCESATEERVYRQNIQERARTEMEDHRTCEAFOL QEEFRLILRNALVHNCNLSQMCVKODELEYRLLEGDVGDIINGFIOKSTGR EPPAPVQNYNDREYVPLIGSPISQSGVIRKPSGLHSGDPTTSPAPASSETLT PTERNELVVASIEVOATQGNLNSADYRALDYTAQNSDELISAGDILAVILLEGE DGMWTVENRGGRFVPSYLEKL"
BASE COUNT 409 a 443 c 526 g 314 t
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 1692
Score: 415.00 Matches: 415
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
US-09-068-377c-1 (1-415) x MM087814 (1-1692)
QY 1 MetMetAlaGlnLeuGlnPheArGAspAlaPheTrpCysArgAspPheThrAlaHisThr 20
Db |||||||
QY 274 ATGATGGCCCGACGCTGCAGATGCCCTTCTGTCAGAGACTTCACGCGCCACACA 333
Db |||||||
QY 21 GLytyrGluValleuLeuGlnArgLeuLeuAspGlyArgLysMetCysLysAspValGlu 40
Db |||||||
QY 334 GGGATGAGGTGCTACTGTCAGAGGCTGTCGACGCGAGAGATGTCAGAGATGTGAG 393
Db |||||||
QY 41 GluLeuLeuArgGlnArgAlaGlnAlaGluGluArgTyrGlyLysGluLeuValGlnIle 60
Db |||||||
QY 394 GAGCTGCTCAACACAGAGGCGCCAGCGAGAGAGTACGGAGAGAGACTGTGCAGATT 453
Db |||||||
QY 61 AlaArgLysAlaGlyGlyGlnThrGluMetAsnSerLeuArgThrSerPheAspSerLeu 80
Db |||||||

Db 454 GCACCAAGGCTGTGGCCACAGACAGATGATTCCTCGAGAGACTCTTGTACTCCCTG 513
QY 81 LysGlnGlnThrGlnAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaLeuArgGlu 100
Db |||||||
Db 514 AAGCAGCAACAGAGAAATGTGGCGACACACATCCAGCTGGCCCTGGCGCTGAG 573
QY 101 GluLeuArgSerLeuGluGluPheArgGluArgGlnLysGlnArgLysLysTyrGlu 120
Db |||||||
Db 574 GAGCTGGGAGACCTCGAGAGATTCGACAGAGACAGAAAGAGCGGAGAAAGTATGAG 633
QY 121 AlaIleMetAspArgValGlnLysSerLysLeuSerLeuTyrLysLysThrMetGluSer 140
Db |||||||
Db 634 GCCATCATGTGACCCCTGTCCAGAAAGACCAAGTTCCTTACAGAGACCATGTGATGCC 693
QY 141 LysLysAlaTyrAspGlnLysCysArgAspAlaAspAlaGlnGluGlnAlaPheGluArg 160
Db |||||||
Db 694 AAGAGGCAATGTGACCGAAGTGCAGGAGATGCAGATATCTGTGAGGCGCTTCGAGCGT 753
QY 161 ValSerAlaAsnGlnLysGlnLysGlnValGlnLysSerGlnAsnLysAlaLysGlnCys 180
Db |||||||
Db 754 GTGAGTGCATATGGCCACACAGACAGATGAGAAAGAGCCAGAAAGCCAGACAGTGC 813
QY 181 LysGlnSerAlaThrGlnAlaGluArgValTyrArgGlnAsnIleGlnLysGluArg 200
Db |||||||
Db 814 AAGAGTCAGCCACAGAGCGAGAAAGAGTACAGCGCAAAATATCGAAACATGTGAGAGA 873
QY 201 AlaArgThrGlnTyrGlnGlnGlnLysArgThrThrCysGluAlaPheGlnLysGln 220
Db |||||||
Db 874 GCGAGGACCGAGTGGAGCAGAGACACCGGACCTACCTGTAGGCTTCGAGTGCAGAG 933
QY 221 PheAspArgLeuThrIleLeuArgAsnAlaLeuTrpValHisCysAsnGlnLysSerMet 240
Db |||||||
Db 934 TTTACCGCGCTACACATCTCCGGAATGCCCTGTGGGTGACACTATACAGACTCTCATG 993
QY 241 GlnCysValLysAspAspGluLeuTyrGlnGluValArgLeuThrLeuGlnLysCysAsp 260
Db |||||||
Db 994 CAGGTGTCAAGATGTAGTACCTTATGAGAAAGTGGCGCTGACCCCTTGAGGCGCTGAT 1053
QY 261 ValGlnGluAspIleAsnGlnPheIleGlnSerLysSerThrGlyArgGluProGlnAla 280
Db |||||||
Db 1054 GTGGAAGGTGCATCATATGCTTCATCCAGTCCAGAGGCACTGGCAGAGAGCCCCAGCT 1113
QY 281 ProValProTyrGlnAsnTyrTyrAspArgGluValThrProLeuIleGlySerProSer 300
Db |||||||
Db 1114 CCGGTGCTTATGACAGACTACTATGACAGGAGGTGATCCCACTGATGGCACCCCTAGC 1173
QY 301 IleGlnProSerCysGlyValIleLysArgPheSerCysLeuLeuHisGlySerProLys 320
Db |||||||
Db 1174 ATCCAGCCCTCTCGGCTGTGATAAAGAGTCTCTGCGCTACATGGAACCTCCCAAG 1233
QY 321 ThrThrProSerAlaProAlaAlaSerThrGluThrLeuThrProThrProGluArgAsn 340
Db |||||||
Db 1234 ACCACACTTCTGCTCTGCTGCTTCCACAGAGACTGTGACTCCACCCCTGAGCGGAAAT 1293
QY 341 GluLeuValTyrAlaSerIleGluValGlnAlaAlaThrGlnGlnLysAsnLeuAsnSerSerAla 360
Db |||||||
Db 1294 GAGTGTGCTACGATCCATCGAAGTGCAGGACGACCCAGGAGAAACCTTAACTATCAGCC 1353
QY 361 GlnAspTyrArgAlaLeuTyrAspTyrThrAlaGlnAsnSerAspGluLeuAspIleSer 380
Db |||||||
Db 1354 CAGCAGTACCGGGGACACTCTACGACTACACAGCAGCAAGATTCGATAGAGCGCAATTTC 1413
QY 381 AlaGlyAspIleLeuAlaValIleLeuGlnGlyGlnAspGlyTrpThrPheValGluArg 400
Db |||||||
Db 1414 GCGGAGACATCTCGGGGTCACTCGGAAGGGGAGAGATGGCTGTGGACTGTGGAGCGG 1473
QY 401 AsnGlnGlnArgGlyPheValProGlySerTyrLeuGlnLysLeu 415
Db |||||||
Db 1474 AACGGAACAGCTGCTTGTTCCTCGGTCTACTTGGAGAGAGCTC 1518
RESULT 2
ARI08626
LOCUS ARI08626 2100 bp DNA linear PAT 14-FEB-2001

DEFINITION Sequence 2 from patent US 6111073.
ACCESSION ARI08626
VERSION ARI08626.1 GI:1282413
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2100)
AUTHORS Lasky, L.A.
TITLE Tyrosine phosphorylated cleavage furrow-associated proteins (PSPiPs)
JOURNAL Patent: US 6111073-A 2 29-AUG-2000;
FEATURES Location/Qualifiers
source 1. 2100
BASE COUNT 548 a 540 c 594 g 418 t
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 2100
Score: 415.00 Matches: 415
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0
US-09-068-377C-1 (1-415) x ARI08626 (1-2100)
QY 1 MetMetAlaGlnLeuGlnPheArgAspAlaPheTrpCysArgAspPheThrAlaHisThr 20
DB 682 ATGATGGCCGCACTGCAGATCCGAGATGCTTCTGTCAGGAGACTTCAAGCCACACA 741
QY 21 G1YTRG1ValLeuLeuGlnPheLeuAspG1YArgLysMetCysAspValGlu 40
DB 742 GGGTGTAGAGTCTCTACTGCAAGAGGCTGTCGACGCGAGAAGATGTGCAAGATGTGAG 801
QY 41 GluLeuLeuArgGlnArgAlaGlnAlaGlnGluArgGluArgGluLysGluLeuValGlnIle 60
DB 802 GAGCTGCTCAGACAGAGGGCCGAGGAGAGAGGTACGAGAGAGAGCTGCTGTCAGATT 861
QY 61 AlaArgLysAlaGlyLysGlnThrGluMetAsnSerLeuArgThrSerPheAspSerLeu 80
DB 862 GCACGCAAGCTGTGTGGCCAGACAGAGATGAATCCCTGAGGACCTCTTGTGACTCCCTG 921
QY 81 LysGlnGlnThrGluAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaLeuArgGlu 100
DB 922 AAGCAGCAACAGAGATGTGGGCACTGCACATCAGCTGCTGCTGCTGCTGCTGCTGCTG 981
QY 101 GluLeuArgSerLeuGlnGluPheArgGluArgGlnLysGlnGlnArgLysLysTYRGLu 120
DB 982 GAGCTGCGGAGCTGTGAGAGATTCGAGAGAGACAGAAAGAGAGCGGAAAGATGTGAG 1041
QY 121 AlaIleMetAspArgValGlnLysSerLysLeuSerLeuTYRTrpLysLysThrMetGluSer 140
DB 1042 GCCATCATGACCGCTGTCCAGAAAGCAAGTGTCTCCCTACAGAAAGACCATGGAGTGC 1101
QY 141 LysLysAlaTYRAspGlnLysCysArgAspAlaAspAspAlaGlnGlnAlaPheGluArg 160
DB 1102 AAGAAAGCATATGACAGAGATGTGCAAGATGTGAGATGTGAGAGCGCTTCGAGCGT 1161
QY 161 ValSerAlaAsnGlyHisGlnLysGlnValGlnLysSerGlnAsnLysAlaLysGlnCys 180
DB 1162 GTGAGTGCCTATGCGCCACAGAAAGCAAGTGTGAGAAAGCAAGCAAGCAAGCAAGTGC 1221
QY 181 LysGluSerAlaThrGlnAlaGluArgValTYRArgGlnAsnIleGluGlnLeuGluArg 200
DB 1222 AAGGAGTCAAGCACAAGAGCAGAAAGATGTACAGCAAAATATTCACAACTGCGAGAGA 1281
QY 201 AlaArgThrGlnTrpGlnGlnGlnHisArgGluThrCysGlnAlaPheGlnLeuGlnGlu 220
DB 1282 GCGAGAGCCAGAGGAGGAGCAGACCGGAGTCACTGTGAGGCTTCAGGTTGCGAGGAG 1341
QY 221 PheAspArgLeuThrIleLeuArgAsnAlaLeuTrpValHisCysAsnGlnLeuSerMet 240

DB 1342 TTGGACCGGCGTACCAATCCCGCCGATGCCCTGTGGGTGAGCTGTAAACAGCTCTCCAG 1401
QY 241 GlnCysValLysAspAspGluLeuTYRTrpGlnGluValArgLeuThrLeuGluGlyCysAsp 260
DB 1402 CAGTGTGTCAAGATATGATGAGCTGTATGAGGAAGTCCGCTGACCTTGAGGGCTGTGAT 1461
QY 261 ValGluGluAspIleAsnGlyPheIleGlnSerLysSerThrGlyArgGluProPheAla 280
DB 1462 GTGGAAGGTGACATCAATGAGCTTCAATCCAGCAAGCAGTGCAGAGAGAGCCCAAGCT 1521
QY 281 ProValProTYRGlInAsnTYRTrpAspArgGluValThrProLeuIleGlySerProSer 300
DB 1522 CCGGTCTCTATATGAGAACTACTATGACAGGAGAGTACCCCATGATGTGGCACCCTAG 1581
QY 301 IleGlnProSerCysGlyValIleLysArgPheSerGlyLeuLeuHisGlySerProLys 320
DB 1582 ATCCAGCCCTCTCCGCTGTGATTAAGAGGTTCTGCGCTGCTACATGGAATCCCAAG 1641
QY 321 ThrThrProSerAlaProAlaAlaSerThrGluThrLeuThrProThrProGluArgAsn 340
DB 1642 ACCACACCTTCTGCTCTGCTGCTTCCACAGAGACTGTACTCCACCTTGAGCGGAAT 1701
QY 341 GluLeuValTYRAlaSerIleGluValGlnAlaThrGlnGlnLysAsnLeuAsnSerAla 360
DB 1702 GAGTGTGCTACAGCATTCATCGAAGTGCAGCGCAGCCAGAGAAACCTTAATCATCAGACC 1761
QY 361 GlnAspTYRArgValAlaLeuTYRAspTYRThrAlaGlnAsnSerAspGluLeuAspIleSer 380
DB 1762 CAGGACTACCGGGCAGCTACGACTACAGCTACAGCTACAGATTCATATGAGTGGACATTTC 1821
QY 381 AlaGlyAspIleLeuAlaValIleLeuGluGlyGlnAspGlyTrpTrpTrpValGluArg 400
DB 1822 GCGGAGACATCTGTGGCGGTATCTGTGAAGGGGAGAGATGCTGTGTGAGCTGTGAGCGG 1881
QY 401 AsnGlyGlnArgGlyPheValProGlySerTYRLeuGlnLysLeu 415
DB 1882 AAGCGAACAGCTGTGCTTGTCCCTGGGTGCTACTTGGAGAAAGCTC 1926
RESULT 3
AC100102 44437 bp DNA linear HTG 22-NOV-2001
LOCUS
DEFINITION Mus musculus clone RP23-38K16, LOW-PASS SEQUENCE SAMPLING.
AC100102
VERSION AC100102.1 GI:17047468
KEYWORDS HTG; HTGS_PHASE0.
SOURCE
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 44437)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Unpublished
JOURNAL 2 (bases 1 to 44437)
AUTHORS
REFERENCES
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B., Brown, A., Camarata, J., Campolano, A., Chang, J., Charazi, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Glinde, S., Gord, S., Goyette, M., Graham, L., Grand-pierre, N., Hages, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lakocque, K., Lamazares, R., Landers, T., Lenockky, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J., Menes, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, U., Peterson, K., Phunhkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,

TITLE
JOURNAL
COMMENT

Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.-J., Young, G.,
Zainoun, J., Zemdek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR

Web site: <http://www-seq.wi.mit.edu>
Contact: submissions@genome.wi.mit.edu

Project Information

Center project name: L13959

Center clone name: 38_K_16

* NOTE: This record contains 56 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1
* 684 783: gap of 100 bp
* 784 1485: contig of 702 bp in length
* 1486 1585: gap of 100 bp
* 1586 2297: contig of 712 bp in length
* 2298 2397: gap of 100 bp
* 2398 3062: contig of 665 bp in length
* 3063 3162: gap of 100 bp
* 3163 3805: contig of 643 bp in length
* 3806 3905: gap of 100 bp
* 3906 4615: contig of 710 bp in length
* 4616 4715: gap of 100 bp
* 4716 5411: contig of 696 bp in length
* 5412 5511: gap of 100 bp
* 5512 6212: contig of 701 bp in length
* 6213 6312: gap of 100 bp
* 6313 7012: contig of 700 bp in length
* 7013 7112: gap of 100 bp
* 7113 7801: contig of 689 bp in length
* 7802 7901: gap of 100 bp
* 7902 8584: contig of 683 bp in length
* 8585 8684: gap of 100 bp
* 8685 9374: contig of 690 bp in length
* 9375 9474: gap of 100 bp
* 9475 10182: contig of 708 bp in length
* 10183 10282: gap of 100 bp
* 10283 10997: contig of 715 bp in length
* 10998 11097: gap of 100 bp
* 11098 11812: contig of 715 bp in length
* 11813 11912: gap of 100 bp
* 11913 12635: contig of 723 bp in length
* 12636 12735: gap of 100 bp
* 12736 13426: contig of 691 bp in length
* 13427 13526: gap of 100 bp
* 13527 14227: contig of 701 bp in length
* 14228 14327: gap of 100 bp
* 14328 15030: contig of 703 bp in length
* 15031 15130: gap of 100 bp
* 15131 15806: contig of 676 bp in length
* 15807 15906: gap of 100 bp
* 15907 16588: contig of 683 bp in length
* 16589 16689: gap of 100 bp
* 16690 17370: contig of 681 bp in length

* 17371 17470: gap of 100 bp
* 17471 18173: contig of 703 bp in length
* 18174 18273: gap of 100 bp
* 18274 18970: contig of 697 bp in length
* 18971 19070: gap of 100 bp
* 19071 19781: contig of 711 bp in length
* 19782 19881: gap of 100 bp
* 19882 20576: contig of 695 bp in length
* 20577 20676: gap of 100 bp
* 20677 21362: contig of 686 bp in length
* 21363 21462: gap of 100 bp
* 21463 22166: contig of 704 bp in length
* 22167 22266: gap of 100 bp
* 22267 22961: contig of 695 bp in length
* 22962 23061: gap of 100 bp
* 23062 23752: contig of 691 bp in length
* 23753 23852: gap of 100 bp
* 23853 24558: contig of 706 bp in length
* 24559 24658: gap of 100 bp
* 24659 25340: contig of 682 bp in length
* 25341 25440: gap of 100 bp
* 25441 26117: contig of 677 bp in length
* 26118 26217: gap of 100 bp
* 26218 26921: contig of 704 bp in length
* 26922 27021: gap of 100 bp
* 27022 27722: contig of 701 bp in length
* 27723 27823: gap of 100 bp
* 27823 28507: contig of 685 bp in length
* 28508 28607: gap of 100 bp
* 28608 29300: contig of 693 bp in length
* 29301 29400: gap of 100 bp
* 29401 30089: contig of 689 bp in length
* 30090 30189: gap of 100 bp
* 30190 30895: contig of 706 bp in length
* 30896 30995: gap of 100 bp
* 30996 31700: contig of 705 bp in length
* 31701 31800: gap of 100 bp
* 31801 32483: contig of 683 bp in length
* 32484 32583: gap of 100 bp
* 32584 33276: contig of 693 bp in length
* 33277 33376: gap of 100 bp
* 33377 34078: contig of 702 bp in length
* 34079 34178: gap of 100 bp
* 34179 34882: contig of 704 bp in length
* 34883 34982: gap of 100 bp
* 34983 35682: contig of 700 bp in length
* 35683 35782: gap of 100 bp
* 35783 36467: contig of 685 bp in length
* 36468 36567: gap of 100 bp
* 36568 37251: contig of 684 bp in length
* 37252 37351: gap of 100 bp
* 37352 38040: contig of 689 bp in length
* 38041 38140: gap of 100 bp
* 38141 38838: contig of 698 bp in length
* 38839 38938: gap of 100 bp
* 38939 39649: contig of 711 bp in length
* 39650 39749: gap of 100 bp
* 39750 40426: contig of 677 bp in length
* 40427 40526: gap of 100 bp
* 40527 41230: contig of 704 bp in length
* 41231 41330: gap of 100 bp
* 41331 42022: contig of 692 bp in length
* 42023 42122: gap of 100 bp
* 42123 42832: contig of 710 bp in length
* 42833 42932: gap of 100 bp
* 42933 43636: contig of 704 bp in length
* 43637 43736: gap of 100 bp
* 43737 44437: contig of 701 bp in length.

FEATURES

Source

1..44437
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="Rp23-38K16"

BASE COUNT 9664 a 9556 c 9429 g 10121 t 5667 others
ORIGIN /clone_lib="RPCI-23 Female Mouse BAC"

Alignment Scores:

Pred. No.:	1.15e-36	Length:	44437
Score:	45.00	Matches:	45
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	10.84%	Indels:	0
DB:	2	Gaps:	0

US-09-068-377C-1 (1-415) x AC101012 (1-44437)

QY 371 Alaglnasneraspcluleuaspiileserlaglvaspilleuualavalileuclu 390
|||||
DB 17205 GCCCAGAAATTCATAGCTGAGACATTTCCGCCGGAGACATCTCGCGGCATCTCGGAA 17264
QY 391 Glygluaspcllytrprrhrvalgluargasnlygluarglyphevalproglyser 410
|||||
DB 17265 GGGGAGAGATGGCTGTGACTGTGAGCGAGACGACGACGCTTGTGCTCGGAGTCG 17324
QY 411 Tyrlengluusleu 415
|||||
DB 17325 TACTTGAGAGAGCTC 17339

RESULT 4
AC123729 72661 bp DNA linear HTG 01-JUN-2002
LOCUS Mus musculus clone RP23-456B7, LOW-PASS SEQUENCE SAMPLING.
AC123729 AC123729.1 GI:21307291
DEFINITION HTG: HTGS PHASE0.
VERSION house mouse.
KEYWORDS Mus musculus
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 72661)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
TITLE Mus musculus, clone RP23-456B7
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 72661)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
TITLE Mus musculus, clone RP23-456B7
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 72661)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
TITLE Mus musculus, clone RP23-456B7
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 72661)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
TITLE Mus musculus, clone RP23-456B7
JOURNAL Unpublished

TITLE Submitted (01-JUN-2002) Whitehead Institute/MIT Center for Genome
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT All repeats were identified using RepeatMasker:
Smiti, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L26489
Center clone name: 456_B_7

NOTE: This record contains 90 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 689: contig of 689 bp in length
* 690 789: gap of 100 bp
* 790 1520: contig of 731 bp in length
* 1521 1620: gap of 100 bp
* 1621 2339: contig of 719 bp in length
* 2340 2439: gap of 100 bp
* 2440 3159: contig of 720 bp in length
* 3160 3259: gap of 100 bp
* 3260 3962: contig of 703 bp in length
* 3963 4062: gap of 100 bp
* 4063 4752: contig of 690 bp in length
* 4753 4852: gap of 100 bp
* 4853 5539: contig of 687 bp in length
* 5540 5639: gap of 100 bp
* 5640 6350: contig of 711 bp in length
* 6351 6450: gap of 100 bp
* 6451 7161: contig of 711 bp in length
* 7162 7261: gap of 100 bp
* 7262 7984: contig of 723 bp in length
* 7985 8084: gap of 100 bp
* 8085 8812: contig of 728 bp in length
* 8813 8912: gap of 100 bp
* 8913 9618: contig of 706 bp in length
* 9619 9718: gap of 100 bp
* 9719 10444: contig of 726 bp in length
* 10445 10544: gap of 100 bp
* 10545 11275: contig of 731 bp in length
* 11276 11375: gap of 100 bp
* 11376 12095: contig of 720 bp in length
* 12096 12195: gap of 100 bp
* 12196 12879: contig of 684 bp in length
* 12880 12979: gap of 100 bp
* 12980 13676: contig of 697 bp in length
* 13677 13776: gap of 100 bp
* 13777 14476: contig of 700 bp in length
* 14477 14576: gap of 100 bp
* 14577 15297: contig of 721 bp in length
* 15298 15397: gap of 100 bp
* 15398 16107: contig of 710 bp in length
* 16108 16207: gap of 100 bp
* 16208 16915: contig of 708 bp in length
* 16916 17015: gap of 100 bp
* 17016 17742: contig of 727 bp in length
* 17743 17842: gap of 100 bp
* 17843 18557: contig of 715 bp in length
* 18558 18657: gap of 100 bp
* 18658 19350: contig of 693 bp in length
* 19351 19450: gap of 100 bp
* 19451 20171: contig of 721 bp in length
* 20172 20271: gap of 100 bp
* 20272 20998: contig of 727 bp in length
* 20999 21098: gap of 100 bp
* 21099 21815: contig of 717 bp in length
* 21816 21915: gap of 100 bp
* 21916 22619: contig of 704 bp in length
* 22620 22719: gap of 100 bp

```

* 22720 23417: contig of 698 bp in length
* 23418 23517: gap of 100 bp
* 23518 24198: contig of 681 bp in length
* 24199 24298: gap of 100 bp
* 24299 25011: contig of 713 bp in length
* 25012 25111: gap of 100 bp
* 25112 25833: contig of 722 bp in length
* 25834 25933: gap of 100 bp
* 25934 26557: contig of 724 bp in length
* 26558 26757: gap of 100 bp
* 26758 27482: contig of 725 bp in length
* 27483 27582: gap of 100 bp
* 27583 28310: contig of 728 bp in length
* 28311 28410: gap of 100 bp
* 28411 29135: contig of 725 bp in length
* 29136 29235: gap of 100 bp
* 29236 29933: contig of 698 bp in length
* 29934 30033: gap of 100 bp
* 30034 30700: contig of 667 bp in length
* 30701 30800: gap of 100 bp
* 30801 31489: contig of 689 bp in length
* 31490 31589: gap of 100 bp
* 31590 32297: contig of 708 bp in length
* 32298 32397: gap of 100 bp
* 32398 33098: contig of 701 bp in length
* 33099 33198: gap of 100 bp
* 33199 33906: contig of 708 bp in length
* 33907 34006: gap of 100 bp
* 34007 34727: contig of 721 bp in length
* 34728 34827: gap of 100 bp
* 34828 35555: contig of 728 bp in length
* 35556 35655: gap of 100 bp
* 35656 36359: contig of 704 bp in length
* 36360 36459: gap of 100 bp
* 36460 37188: contig of 729 bp in length
* 37189 37288: gap of 100 bp
* 37289 38014: contig of 726 bp in length
* 38015 38114: gap of 100 bp
* 38115 38835: contig of 721 bp in length
* 38836 38935: gap of 100 bp
* 38936 39625: contig of 690 bp in length
* 39626 39725: gap of 100 bp
* 39726 40432: contig of 707 bp in length
* 40433 40532: gap of 100 bp
* 40533 41231: contig of 699 bp in length
* 41232 41331: gap of 100 bp
* 41332 42030: contig of 699 bp in length
* 42031 42130: gap of 100 bp
* 42131 42828: contig of 689 bp in length
* 42829 42929: gap of 100 bp
* 42930 43637: contig of 708 bp in length
* 43638 43737: gap of 100 bp
* 43738 44463: contig of 726 bp in length
* 44464 44563: gap of 100 bp
* 44564 45288: contig of 725 bp in length
* 45289 45388: gap of 100 bp
* 45389 46110: contig of 722 bp in length
* 46111 46210: gap of 100 bp
* 46211 46926: contig of 716 bp in length
* 46927 47026: gap of 100 bp
* 47027 47724: contig of 698 bp in length
* 47725 47824: gap of 100 bp
* 47825 48523: contig of 699 bp in length
* 48524 48623: gap of 100 bp
* 48624 49312: contig of 689 bp in length
* 49313 49412: gap of 100 bp
* 49413 50103: contig of 691 bp in length
* 50104 50203: gap of 100 bp
* 50204 50915: contig of 712 bp in length
* 50916 51015: gap of 100 bp
* 51016 51731: contig of 716 bp in length
* 51732 51831: gap of 100 bp
* 51832 52552: contig of 721 bp in length

```

```

* 52553 52652: gap of 100 bp
* 52653 53383: contig of 731 bp in length
* 53384 53483: gap of 100 bp
* 53484 54185: contig of 702 bp in length
* 54186 54285: gap of 100 bp
* 54286 55013: contig of 728 bp in length
* 55014 55113: gap of 100 bp
* 55114 55838: contig of 725 bp in length
* 55839 55938: gap of 100 bp

```

Alignment Scores:

```

Pred. No.: 1,81e-36 Length: 72661
Score: 45.00 Matches: 45
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.84% Indels: 0
DB: 2 Gaps: 0

```

US-09-068-377C-1 (1-415) x AC123729 (1-72661)

OY 371 AlaglnasnsrAspgIulnAspIleSeRAlagIyAspIleuAlaValIleuGlu 390

DB 48675 GCCCAGATTCGTGATGAGCTGAGACATTTCCGCGGAGACATCTGGCGGTATCTGGGA 48734

OY 391 GlyIuAspgIyTPTPTPTnValGluARgAsngIyGlnARgIyPheValProGlySer 410

DB 48735 GGGGAGATGAGCTGCTGTGAGCTGAGCGAGCGAGCAACAGTGGCTTTGTCCCTGGGTG 48794

OY 411 TyrlenuIuIysIleu 415

DB 48795 TACTTGGAGAGCTC 48809

RESULT 5

AC116689/c

Mus musculus clone RP24-20IN16, WORKING DRAFT SEQUENCE, 12 ordered

DEFINITION

LOCUS

AC116689

VERSION

AC116689.3

KEYWORDS

HTG: HTGS_PHAZE2; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE

ORGANISM

Mus musculus

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

Submitted (02-APR-2002) Whitehead Institute/MIT Center for Genome

REFERENCE
AUTHORS
1 (bases 1 to 173635)
Munty D.M., Adams C., Adio-Oduola B., Ali-Osman F.R., Allen C.,
Albrooks S.L., Amaraltinge H.C., Are J.R., Ayele M., Banks T.,
Barbata J., Benton J., Blinze K., Blankenburg K., Bonin D.,
Bouck J., Bowler S., Bivela M., Brown E., Brown M., Bryan N.P.,
Bunay C., Burck P., Burkett C., Burrell K.L., Byrd N.C.,
Carron T.F., Carter M., Cavazos S.R., Chacko J., Chavez D.,
Chen G., Chen R., Chen Z., Chowdhury I., Christopoulos C.,
Cleveland C.D., Cox C., Coyle M.D., Dathore S.R., David R.,
Davila M.L., Davis C., Davy Carroll L., Dederich D.A.,
DeLaney K.R., Delgado O., Dem A.L., Ding Y., Dinh H.H.,
Douthwaite K.J., Draper H., Dugan-Rocha S., Durbin K.J.,
Eamhart C., Edgar D., Edwards C.C., Elhaj C., Escotto M.,
Falls T., Ferraguto D., Flagg N., Ford J., Foster P., Frantz P.,
Gabali A., Gao J., Garcia A., Garner T., Garza N., Gill R.,
Gorell J.H., Guevara W., Gunaratne P., Hale S., Hamilton K.,
Harris C., Harris K., Hart M., Havlak P., Hawes A., Hernandez J.,
Hernandez O., Hodgson A., Hogues M., Holloway C., Hollins B.,
Homs F., Howard S., Huber J., Hulyk S., Hume J., Jackson L.E.,
Jacobson B., Jia Y., Johnson R., Jolivet S., Joudah S.,
Karlsen E., Kelly S., Khan U., King L., Korvah J., Kovar C.,
Kratovic J., Kureshi A., Landry N., Leal B., Lewis L.C., Lewis L.,
Li J., Li Z., Licharge O., Lieu C., Liu J., Liu W., Louised H.,
Locado R.J., Lu X., Lucier A., Lucier R., Luna R., Ma J.,
Madesmart M., Mapa P., Martin R., Martindale A., Martinez E.,
Masse E., Mahoney E., McLeod M.P., Meador M., Mel G., Metzger M.,
Miner G., Miner Z., Mitchell T., Monabhat K., Morgan M., Morris S.,
Moser M., Neal D., Newton J., Newton N., Nguyen A., Nguyen N.,
Nguyen N., Nickerson E., Nwokenkwo S., Ogun M., Okunolu G.,
Oragunye N., Oviedo R., Pace A., Payton B., Peery J., Perez L.,
Peterson L., Pickens R., Primus E., Pu L., Quiles M., Ren Y.,
Rives M., Rojas A., Rojibokan I., Rolfe M., Ruiz S., Saverly G.,
Schier S., Scott G., Shen H., Shoshitari N., Sisson I.,
Soderren E., Sonalke T., Sparks A., Stanley H., Stone H.,
Sutton A., Swalek A., Taber P., Tamerisa A., Tamerisa K., Tang H.,
Tansey J., Taylor C., Taylor T., Telford B., Thomas R., Thomas S.,
Usmani K., Vasquez L., Vera V., Villalobos D., Vinson R., Wang Q.,
Wang S., Ward-Moore S., Warren R., Washington C., Watlington S.,
Williams G., Williamson A., Wlezyk R., Woodson S., Worley K.,
Wu C., Wu Y., Wu Y.F., Zhou J., Zorrilla S., Nelson D.,
Weinstock G. and Gibbs R.
Direct Submission
Unpublished
2 (bases 1 to 173635)
Worley K.C.
Direct Submission
Submitted (29-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 173635)
Worley K.C.
Direct Submission
Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GXGD
Center clone name: CH230-430H3
Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye 1000 of reads
Assembly program: Phrap, version 0.990329
Consensus quality: 143651 bases at least Q40
Consensus quality: 149971 bases at least Q30
Consensus quality: 153974 bases at least Q20
* NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 42 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1473: contig of 1473 bp in length
1 1474
1 1573: gap of unknown length
1 1574
1 2998: contig of 1425 bp in length
1 2999
1 3098: gap of unknown length
1 3099
1 4444: contig of 1346 bp in length
1 4445
1 4544: gap of unknown length
1 4545
1 5810: contig of 1266 bp in length
1 5811
1 5910: gap of unknown length
1 5911
1 7362: contig of 1452 bp in length
1 7363
1 7462: gap of unknown length
1 7463
1 9579: contig of 2117 bp in length
1 9580
1 9679: gap of unknown length
1 10981: contig of 1302 bp in length
1 10982
1 11081: gap of unknown length
1 11082
1 12641: contig of 1560 bp in length
1 12642
1 12741: gap of unknown length
1 12742
1 14658: contig of 1917 bp in length
1 14659
1 14758: gap of unknown length
1 14759
1 17461: contig of 2703 bp in length
1 17462
1 17561: gap of unknown length
1 17562
1 19150: contig of 1589 bp in length
1 19151
1 19250: gap of unknown length
1 19251
1 21673: contig of 2423 bp in length
1 21674
1 21774: gap of unknown length
1 21775
1 24476: contig of 2703 bp in length
1 24477
1 24576: gap of unknown length
1 24577
1 27228: contig of 2652 bp in length
1 27229
1 27328: gap of unknown length
1 27329
1 31213: contig of 3885 bp in length
1 31214
1 31313: gap of unknown length
1 31314
1 33332: contig of 2019 bp in length
1 33333
1 33433: gap of unknown length
1 33434
1 35180: contig of 1748 bp in length
1 35181
1 35280: gap of unknown length
1 35281
1 38329: contig of 3049 bp in length
1 38330
1 38429: gap of unknown length
1 38430
1 41439: contig of 3010 bp in length
1 41440
1 41539: gap of unknown length
1 41540
1 44542: contig of 3003 bp in length
1 44543
1 44642: gap of unknown length
1 44643
1 47989: contig of 3347 bp in length
1 47990
1 48089: gap of unknown length
1 48090
1 51923: contig of 3834 bp in length
1 51924
1 52023: gap of unknown length
1 52024
1 56630: contig of 4607 bp in length
1 56631
1 56730: gap of unknown length
1 56731
1 59866: contig of 3136 bp in length
1 59867
1 59967: gap of unknown length
1 59968
1 62219: contig of 2253 bp in length
1 62220
1 62319: gap of unknown length
1 62320
1 65464: contig of 3145 bp in length
1 65465
1 65564: gap of unknown length
1 65565
1 69463: contig of 3899 bp in length
1 69464
1 74055: gap of unknown length
1 74056
1 74155: contig of 4492 bp in length
1 74156
1 77949: gap of unknown length
1 77950
1 78049: contig of 3794 bp in length
1 78050
1 81228: gap of unknown length
1 81229
1 81328: gap of unknown length
1 81329
1 84778: contig of 3450 bp in length
1 84779
1 84878: gap of unknown length
1 84879
1 89049: contig of 4171 bp in length
1 89149: gap of unknown length
1 89050

89150 93815: contig of 4666 bp in length
* 93816 93915: gap of unknown length
* 93916 100515: contig of 6600 bp in length
* 100516 100615: gap of unknown length
* 100616 105660: contig of 5045 bp in length
* 105661 105760: gap of unknown length
* 105761 111730: contig of 5970 bp in length
* 111731 111830: gap of unknown length
* 111831 117830: contig of 6000 bp in length
* 117831 117930: gap of unknown length
* 117931 125300: contig of 7370 bp in length
* 125301 125400: gap of unknown length
* 125401 134911: contig of 9511 bp in length
* 134912 135011: gap of unknown length
* 135012 143270: contig of 8259 bp in length
* 143271 143370: gap of unknown length
* 143371 151512: contig of 8142 bp in length
* 151513 151612: gap of unknown length
* 151613 173635: contig of 22023 bp in length.
Location/Qualifiers
1. 173635
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-430H3"
BASE COUNT 43436 a 39410 c 39651 g 45570 t 5568 others
ORIGIN
Alignment Scores:
Pred. NO.: 4.03e-36 Length: 173635
Score: 45.00 Matches: 45
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.84% Indels: 0
DB: 2 Gaps: 0
US-09-068-377c-1 (1-415) x AC123432 (1-173635)
QY 371 Alaglnasnsersapgluleuaspiileseralaglyaspilaleualavallileucln 390
|||||
Db 163306 GCCCAGATTCTGATGACGTGACATTTCTCGGAGACATCTCGCATCTGCAA 163247
QY 391 G1YGLnAspGlyTTPTPPhValGluArGAsnGlyGlnArGlyPheValProGlySer 410
|||||
Db 163246 GGGGAGGATGCTGCTGAGCTGAGCAACGAGCAACAGCTGCTTGTCTCCGTGTTGC 163187
QY 411 TyrlenculuySteu 415
|||||
Db 163186 TACTTGAGAGAGCTC 163172
RESULT 7
AC123482/c 200703 bp DNA linear HTG 22-JUL-2002
LOCUS Rattus norvegicus clone CH230-61L123, *** SEQUENCING IN PROGRESS
DEFINITION *** 62 unordered pieces.
AC123482
AC123482.2 GI:21908373
VERSION HTG: HTGS PHASE1.
KEYWORDS Rattus norvegicus.
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 200703)
REFERENCE
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Aisbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbarta,J., Benton,J., Bimarge,K., Blankenburg,K., Bonnin,D.,
Bouch,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhaey,C., Burich,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carion,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,

Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,U., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Hollway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulvik,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudh,S.,
Karlssohn,E., Kelly,S., Khan,U., King,L., Korvan,T., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H.,
Lorado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Mosier,M., Neal,D., Newton,N., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokwkw,S., Ogub,M., Okunolu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Qules,M., Ren,Y.,
Rivers,M., Rojas,A., Rojubenkan,I., Rolfe,M., Ruiz,S., Severy,G.,
Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczek,R., Woodson,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhu,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Glbbs,R.
Direct Submission
Unpublished
2 (bases 1 to 200703)
Worley,K.C.
Direct Submission
Submitted (29-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 200703)
Worley,K.C.
Direct Submission
Submitted (22-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 19, 2002 this sequence version replaced gi:21240424.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center Project name: GX00
Center Clone name: CH230-61L123
----- Summary Statistics
Sequencing vector: Plasmid:
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 150694 bases at least Q40
Consensus quality: 155595 bases at least Q30
Consensus quality: 160071 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 62 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1003: contig of 1003 bp in length
* 1004 1103: gap of unknown length

```

1104      2208: contig of 1105 bp in length
2209      2308: gap of unknown length
2309      3563: contig of 1255 bp in length
3564      4783: gap of unknown length
3664      4783: contig of 1120 bp in length
4784      4883: gap of unknown length
4884      6071: contig of 1188 bp in length
6072      6171: gap of unknown length
6172      7779: contig of 1608 bp in length
7780      7879: gap of unknown length
7880      9371: contig of 1492 bp in length
9372      9471: gap of unknown length
9472      10504: contig of 1033 bp in length
10505      10604: gap of unknown length
10605      12352: contig of 1748 bp in length
12353      12452: gap of unknown length
12453      13742: contig of 1290 bp in length
13743      13842: gap of unknown length
13843      15465: contig of 1623 bp in length
15466      15565: gap of unknown length
15566      16746: contig of 1181 bp in length
16747      16846: gap of unknown length
16847      17935: contig of 1089 bp in length
17936      18035: gap of unknown length
18036      19676: contig of 1641 bp in length
19677      19776: gap of unknown length
19777      21042: contig of 1266 bp in length
21043      21142: gap of unknown length
21143      22151: contig of 1009 bp in length
22152      22251: gap of unknown length
22252      23631: contig of 1280 bp in length
23632      24981: contig of 1350 bp in length
24982      25081: gap of unknown length
25082      27185: contig of 2104 bp in length
27186      27285: gap of unknown length
27285      30088: contig of 2803 bp in length
30089      30188: gap of unknown length
30189      31697: contig of 1509 bp in length
31698      31797: gap of unknown length
31798      32843: contig of 1046 bp in length
32844      32943: gap of unknown length
32944      34632: contig of 1689 bp in length
34633      34732: gap of unknown length
34733      37595: contig of 2863 bp in length
37596      37695: gap of unknown length
37696      39605: contig of 1910 bp in length
39606      39705: gap of unknown length
39706      43903: contig of 4198 bp in length
43904      44003: gap of unknown length
44004      46645: contig of 2642 bp in length
46646      46745: gap of unknown length
46746      49210: contig of 2465 bp in length
49211      49310: gap of unknown length
49311      50878: contig of 1568 bp in length
50879      50978: gap of unknown length
50979      52820: contig of 1842 bp in length
52821      52920: gap of unknown length
52921      54162: contig of 1242 bp in length
54163      54262: gap of unknown length
54263      56801: contig of 2539 bp in length
56802      56901: gap of unknown length
56902      60486: contig of 3585 bp in length
60487      60586: gap of unknown length
60587      63702: contig of 3116 bp in length
63703      63802: gap of unknown length
63803      67541: contig of 3739 bp in length
67542      67641: gap of unknown length
67642      71052: contig of 3411 bp in length
71053      71152: gap of unknown length
71153      73627: contig of 2475 bp in length
73628      73727: gap of unknown length
73728      77051: contig of 3324 bp in length

```

```

77052      77151: gap of unknown length
77152      79762: contig of 2611 bp in length
79763      79862: gap of unknown length
79863      82516: contig of 2654 bp in length
82517      82617: gap of unknown length
82617      85158: contig of 2542 bp in length
85159      85259: gap of unknown length
85259      88880: contig of 3622 bp in length
88881      88980: gap of unknown length
88981      91550: contig of 2570 bp in length
91551      91650: gap of unknown length
91651      95912: contig of 4262 bp in length
95913      96012: gap of unknown length
96013      98576: contig of 2564 bp in length
98577      98676: gap of unknown length
98677      102039: contig of 3363 bp in length
102040      102139: gap of unknown length
102140      105631: contig of 3492 bp in length
105632      105731: gap of unknown length
105732      109208: contig of 3477 bp in length
109209      109308: gap of unknown length
109309      113534: contig of 4225 bp in length
113534      113633: gap of unknown length
113633      120988: contig of 7355 bp in length
120989      121088: gap of unknown length
121089      126035: contig of 4946 bp in length
126035      126134: gap of unknown length
126135      133969: contig of 7835 bp in length
133970      134069: gap of unknown length
134070      138869: contig of 4900 bp in length
138970      139069: gap of unknown length
139070      143826: contig of 4757 bp in length

```

Alignment Scores:

```

Pred. No.: 4 6e-36 Length: 200703
Score: 45.00 Matches: 45
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.84% Indels: 0
DB: 2 Gaps: 0

```

US-09-068-377c-1 (1-415) x AC1123482 (1-200703)

```

QY      371 AlagInasSerFasgLUleuAspIleSerAlagIAspIleuAlaValIleuLgU 390
      |||||||
DB      138202 GCCCAGAAATTCGATGAGACTGGACATTTCTGGGGAGACACCTCGCATCTGGAA 138143
      |||||||
QY      391 GtlyuAspGlyTTPTrThValGluArgAsnGlyInArgGlyPheValProGlySer 410
      |||||||
DB      138142 GGGGAGATGCTGCTGGAGACTGTGCAACGACAAACGTCGCTTGTCCCTGTCG 138083
      |||||||
QY      411 TyrlengLUySLeu 415
      |||||||
DB      138082 TACTTGGAGAGAGCTC 138068

```

RESULT 8

AC101133/C 57518 bp DNA linear. HTG 23-NOV-2001

LOCUS Mus musculus clone Rp23-170K14, LOW-PASS SEQUENCE SAMPLING.

DEFINITION AC101133.1 GI:17059907

ACCESSION AC101133.1 GI:17059907

VERSION AC101133.1 GI:17059907

KEYWORDS HTG; HTGS; PHASEO.

SOURCE Mus musculus.

ORGANISM Mus musculus.

REFERENCE 1 (bases 1 to 57518)

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE Mus musculus clone Rp23-170K14

JOURNAL Unpublished

2 (bases 1 to 57518)

REFERENCE Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,

Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B.,

TITLE
JOURNAL
COMMENT

Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Chopel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Ferro, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Garryna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamet, A., Karatas, A., Kells, C., Labocque, K.,
Lamzars, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J.,
Menees, L., Mhova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Piere, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,
Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L16002

Center clone name: 170_K_14

* NOTE: This record contains 73 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will

* be preserved.
* 1
* 711 810: contig of 710 bp in length
* 811 1510: contig of 700 bp in length
* 1511 1610: gap of 100 bp
* 1611 2307: contig of 697 bp in length
* 2308 2407: gap of 100 bp
* 2408 3052: contig of 685 bp in length
* 3093 3192: gap of 100 bp
* 3193 3863: contig of 671 bp in length
* 3864 3963: gap of 100 bp
* 3964 4627: contig of 664 bp in length
* 4628 4727: gap of 100 bp
* 4728 5408: contig of 681 bp in length
* 5409 5508: gap of 100 bp
* 5509 6186: contig of 678 bp in length
* 6187 6286: gap of 100 bp
* 6287 6960: contig of 674 bp in length
* 6961 7060: gap of 100 bp
* 7061 7742: contig of 682 bp in length
* 7743 7842: gap of 100 bp
* 7843 8511: contig of 669 bp in length
* 8512 8611: gap of 100 bp
* 8612 9318: contig of 707 bp in length
* 9319 9418: gap of 100 bp
* 9419 10118: contig of 700 bp in length
* 10119 10218: gap of 100 bp
* 10219 10927: contig of 709 bp in length

* 10928 11027: gap of 100 bp
* 11028 11739: contig of 712 bp in length
* 11740 11839: gap of 100 bp
* 11840 12539: contig of 700 bp in length
* 12540 12639: gap of 100 bp
* 12640 13337: contig of 698 bp in length
* 13338 13437: gap of 100 bp
* 13438 14146: contig of 709 bp in length
* 14147 14246: gap of 100 bp
* 14247 14942: contig of 696 bp in length
* 14943 15042: gap of 100 bp
* 15043 15732: contig of 690 bp in length
* 15733 15832: gap of 100 bp
* 15833 16502: contig of 670 bp in length
* 16503 16602: gap of 100 bp
* 16603 17253: contig of 651 bp in length
* 17254 17353: gap of 100 bp
* 17354 18025: contig of 672 bp in length
* 18026 18125: gap of 100 bp
* 18126 18801: contig of 676 bp in length
* 18802 18901: gap of 100 bp
* 18902 19608: contig of 707 bp in length
* 19609 19708: gap of 100 bp
* 19709 20381: contig of 673 bp in length
* 20382 20481: gap of 100 bp
* 20482 21186: contig of 705 bp in length
* 21187 21286: gap of 100 bp
* 21287 22001: contig of 715 bp in length
* 22002 22101: gap of 100 bp
* 22102 22790: contig of 689 bp in length
* 22791 22890: gap of 100 bp
* 22891 23598: contig of 708 bp in length
* 23599 23698: gap of 100 bp
* 23699 24405: contig of 707 bp in length
* 24406 24505: gap of 100 bp
* 24506 25180: contig of 675 bp in length
* 25181 25280: gap of 100 bp
* 25281 25979: contig of 699 bp in length
* 25980 26079: gap of 100 bp
* 26080 26763: contig of 684 bp in length
* 26764 26863: gap of 100 bp
* 26864 27565: contig of 702 bp in length
* 27566 27665: gap of 100 bp
* 27666 28351: contig of 686 bp in length
* 28352 28451: gap of 100 bp
* 28452 29151: contig of 700 bp in length
* 29152 29251: gap of 100 bp
* 29252 29935: contig of 684 bp in length
* 29936 30035: gap of 100 bp
* 30036 30734: contig of 699 bp in length
* 30735 30834: gap of 100 bp
* 30835 31511: contig of 677 bp in length
* 31512 31611: gap of 100 bp
* 31612 32305: contig of 694 bp in length
* 32306 32405: gap of 100 bp
* 32406 33072: contig of 667 bp in length
* 33073 33172: gap of 100 bp
* 33173 33842: contig of 670 bp in length
* 33843 33942: gap of 100 bp
* 33943 34634: contig of 692 bp in length
* 34635 34734: gap of 100 bp
* 34735 35432: contig of 698 bp in length
* 35433 35532: gap of 100 bp
* 35533 36231: contig of 699 bp in length
* 36232 36331: gap of 100 bp
* 36332 37029: contig of 698 bp in length
* 37030 37129: gap of 100 bp
* 37130 37831: contig of 702 bp in length
* 37832 37931: gap of 100 bp
* 37932 38597: contig of 666 bp in length
* 38598 38697: gap of 100 bp
* 38698 39363: contig of 666 bp in length
* 39364 39463: gap of 100 bp

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1473: contig of 1473 bp in length
1474 1573: gap of unknown length
1574 2998: contig of 1425 bp in length
2999 3098: gap of unknown length
3099 4444: contig of 1366 bp in length
4445 4545: gap of unknown length
4545 5810: contig of 1266 bp in length
5811 5911: gap of unknown length
5911 7363: contig of 1452 bp in length
7363 7463: gap of unknown length
7463 9579: contig of 2117 bp in length
9580 9679: gap of unknown length
9680 10981: contig of 1302 bp in length
10982 11081: gap of unknown length
11082 12641: contig of 1560 bp in length
12642 12741: gap of unknown length
12742 14658: contig of 1917 bp in length
14659 14758: gap of unknown length
14759 17461: contig of 2703 bp in length
17462 17561: gap of unknown length
17562 19150: contig of 1589 bp in length
19151 19250: gap of unknown length
19251 21673: contig of 2423 bp in length
21674 21773: gap of unknown length
21774 24476: contig of 2703 bp in length
24477 24576: gap of unknown length
24577 27228: contig of 2652 bp in length
27229 27328: gap of unknown length
27329 31213: contig of 3885 bp in length
31214 33333: gap of unknown length
33333 33432: gap of unknown length
33433 35180: contig of 1748 bp in length
35181 35280: gap of unknown length
35281 38329: contig of 3049 bp in length
38330 38429: gap of unknown length
38430 41439: contig of 3010 bp in length
41440 41539: gap of unknown length
41540 44542: contig of 3003 bp in length
44543 44642: gap of unknown length
44643 47989: contig of 3347 bp in length
47990 48089: gap of unknown length
48090 51923: contig of 3834 bp in length
51924 52023: gap of unknown length
52024 56630: contig of 4607 bp in length
56631 56730: gap of unknown length
56731 59866: contig of 316 bp in length
59867 59966: gap of unknown length
59967 62219: contig of 2253 bp in length
62220 62319: gap of unknown length
62320 65464: contig of 3145 bp in length
65465 65564: gap of unknown length
65565 69463: contig of 3899 bp in length
69464 69563: gap of unknown length
69564 74055: contig of 4492 bp in length
74056 74155: gap of unknown length
74156 77949: contig of 3794 bp in length
77950 81228: contig of 3179 bp in length
81229 81328: gap of unknown length
81329 84778: contig of 3450 bp in length
84779 84878: gap of unknown length
84879 89049: contig of 4171 bp in length
89050 89149: gap of unknown length
89150 93815: contig of 4666 bp in length
93816 93915: gap of unknown length
93916 100515: contig of 6600 bp in length
100515 100615: gap of unknown length

100616 105660: contig of 5045 bp in length
105661 105760: gap of unknown length
105761 111730: contig of 5970 bp in length
111731 111830: gap of unknown length
111831 117831: contig of 6000 bp in length
117832 117930: gap of unknown length
117931 125300: contig of 7370 bp in length
125301 125400: gap of unknown length
125401 134911: contig of 9511 bp in length
134912 135011: gap of unknown length
135012 143270: contig of 8259 bp in length
143271 143370: gap of unknown length
143371 151512: contig of 8142 bp in length
151513 151612: gap of unknown length
151613 173635: contig of 22023 bp in length.

FEATURES
source location/Qualifiers
1..173635
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-430H3"

BASE COUNT 43436 a 39410 c 39651 g 45570 t 5568 others

ORIGIN

Alignment Scores:
Pred. No.: 2,46e-24 Length: 173635
Score: 34.00 Matches: 34
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.19% Indels: 0
DB: 2 Gaps: 0

US-09-068-377C-1 (1-415) x AC123432 (1-173635)

QY 13 CysArgAspPheThrAlaHisThrGlyTyrGluValLeuLeuLeuGlnArgLeuLeuAspGly 32
|||||
DB 29136 TCGAGGAGCTTCACGCGCCACACAGGCGTATGAGTACTCTCGACAGGCTGTGACGCG 29195
|||||

QY 33 ArgGlyMetCysLysAspValGluGluLeuLeuArgGlnArg 46
|||||

DB 29196 AGGAGATGTGCAGAGATGTGCAGAGCGTCGTCACGACACAGG 29237
|||||

RESULT 10
AC123482
LOCUS
DEFINITION
Rattus norvegicus clone CH230-61L23, *** SEQUENCING IN PROGRESS
AC123482 200703 bp DNA linear HTG 22-JUL-2002
AC123482
VERSION
KEYWORDS
HTG: HTGS_PHASE1.
SOURCE
Rattus norvegicus.
ORGANISM
Rattus norvegicus.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 200703)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alpbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayala,M., Banks,T.,
Barbarta,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dalhorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Doutwaite,K.J., Draper,H., Dugan-Rocha,S., Durkin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgeson,A., Hogues,M., Holloway,C., Hollins,B.,
Homsl,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,

[illegible]

* 88881 88980: gap of unknown length
 * 88981 91550: contig of 2570 bp in length
 * 91551 91650: gap of unknown length
 * 91651 95912: contig of 4262 bp in length
 * 95913 96012: gap of unknown length
 * 96013 98576: contig of 2564 bp in length
 * 98577 98676: gap of unknown length
 * 98677 102039: contig of 3363 bp in length
 * 102040 102139: gap of unknown length
 * 102140 105631: contig of 3492 bp in length
 * 105632 105731: gap of unknown length
 * 105732 109208: contig of 3477 bp in length
 * 109209 109308: gap of unknown length
 * 109309 113533: contig of 4225 bp in length
 * 113534 113633: gap of unknown length
 * 113634 120988: contig of 7355 bp in length
 * 120989 121088: gap of unknown length
 * 121089 126034: contig of 4946 bp in length
 * 126035 126135: gap of unknown length
 * 126135 133969: contig of 7835 bp in length
 * 133970 134069: gap of unknown length
 * 134070 138969: contig of 4900 bp in length
 * 138970 139069: gap of unknown length
 * 139070 143826: contig of 4757 bp in length

Alignment Scores:

Pred. No.: 2, 81e-24 Length: 200703
 Score: 34.00 Matches: 34
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.19% Indels: 0
 DB: 2 Gaps: 0

US-09-068-377c-1 (1-415) x AC123482 (1-200703)

QY 13 CysAtgAspPheTherAlaHisThrGlyTYRGlValLeuGlnArgLeuLeuAspGly 32
 DB 173389 TCAGAGGACTTACGCGCCACAGAGGATGAGGTACTCTCGACAGAGCTGCTGAGAGCGC 173448

QY 33 ArgGlyMetCysLysAspValGluGluLeuLeuArgGlnArg 46
 DB 173449 AGGAGATGTGCAAGATGTGGAGAGCTGCTCAGACAGAGC 173490

RESULT 11
 AC100102/c 44437 bp DNA linear HTG 22-NOV-2001
 LOCUS Mus musculus clone RP23-38K16, LOW-PASS SEQUENCE SAMPLING.
 AC100102
 AC100102.1 GI:17047468
 VERSION HTG: HTGS_PHASE0.
 KEYWORDS Mus musculus.
 SOURCE Mus musculus.
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 44437)
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 TITLE Mus musculus, clone RP23-38K16
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 44437)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campione, A., Chang, J., Chazaro, B., Choquet, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Ferro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, N., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hages, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Labocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., MacDonald, P., Major, J., Margulis, N., Matthews, C., McCarthy, M., McKean, P., McKernan, K., McPheters, R., Meldrum, J., Menus, L., Minova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,

TITLE
JOURNAL
COMMENT

Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, M., Riback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: L13959
 Center clone name: 38_K_16

* NOTE: This record contains 56 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will

1 683: contig of 683 bp in length
 * 684 783: gap of 100 bp
 * 784 1485: contig of 702 bp in length
 * 1486 1585: gap of 100 bp
 * 1586 2297: contig of 712 bp in length
 * 2298 2397: gap of 100 bp
 * 2398 3062: contig of 665 bp in length
 * 3063 3162: gap of 100 bp
 * 3163 3805: contig of 643 bp in length
 * 3806 3905: gap of 100 bp
 * 3906 4615: contig of 710 bp in length
 * 4616 4715: gap of 100 bp
 * 4716 5411: contig of 696 bp in length
 * 5412 5511: gap of 100 bp
 * 5512 6212: contig of 701 bp in length
 * 6213 6312: gap of 100 bp
 * 6313 7012: contig of 700 bp in length
 * 7013 7112: gap of 100 bp
 * 7113 7801: contig of 689 bp in length
 * 7802 7901: gap of 100 bp
 * 7902 8584: contig of 683 bp in length
 * 8585 8684: gap of 100 bp
 * 8685 9374: contig of 690 bp in length
 * 9375 9474: gap of 100 bp
 * 9475 10182: contig of 708 bp in length
 * 10183 10282: gap of 100 bp
 * 10283 10997: contig of 715 bp in length
 * 10998 11097: gap of 100 bp
 * 11098 11812: contig of 715 bp in length
 * 11813 11912: gap of 100 bp
 * 11913 12635: contig of 723 bp in length
 * 12636 12735: gap of 100 bp
 * 12736 13426: contig of 691 bp in length
 * 13427 13526: gap of 100 bp
 * 13527 14227: contig of 701 bp in length
 * 14228 14327: gap of 100 bp
 * 14328 15030: contig of 703 bp in length
 * 15031 15130: gap of 100 bp
 * 15131 15806: contig of 676 bp in length

```

* 15807 15906: gap of 100 bp
* 15907 16588: contig of 683 bp in length
* 16590 16688: gap of 100 bp in length
* 16590 17370: contig of 681 bp in length
* 17371 17470: gap of 100 bp in length
* 17471 18173: contig of 703 bp in length
* 18174 18273: gap of 100 bp in length
* 18274 18970: contig of 697 bp in length
* 18971 19070: gap of 100 bp in length
* 19071 19781: contig of 711 bp in length
* 19782 19881: gap of 100 bp in length
* 19882 20576: contig of 695 bp in length
* 20577 20676: gap of 100 bp in length
* 20677 21362: contig of 686 bp in length
* 21363 21462: gap of 100 bp in length
* 21463 22166: contig of 704 bp in length
* 22167 22266: gap of 100 bp in length
* 22267 22961: contig of 695 bp in length
* 22962 23061: gap of 100 bp in length
* 23062 23752: contig of 691 bp in length
* 23753 23852: gap of 100 bp in length
* 23853 24558: contig of 706 bp in length
* 24559 25340: contig of 682 bp in length
* 25341 25440: gap of 100 bp in length
* 25441 26117: contig of 677 bp in length
* 26118 26217: gap of 100 bp in length
* 26218 26921: contig of 704 bp in length
* 26922 27021: gap of 100 bp in length
* 27022 27722: contig of 701 bp in length
* 27723 27822: gap of 100 bp in length
* 27823 28507: contig of 685 bp in length
* 28508 28607: gap of 100 bp in length
* 28608 29300: contig of 693 bp in length
* 29301 29400: gap of 100 bp in length
* 29401 30089: contig of 689 bp in length
* 30090 30189: gap of 100 bp in length
* 30190 30895: contig of 706 bp in length
* 30896 30995: gap of 100 bp in length
* 30996 31700: contig of 705 bp in length
* 31701 31800: gap of 100 bp in length
* 31801 32483: contig of 683 bp in length
* 32484 32583: gap of 100 bp in length
* 32584 33276: contig of 693 bp in length
* 33277 33376: gap of 100 bp in length
* 33377 34078: contig of 702 bp in length
* 34079 34178: gap of 100 bp in length
* 34179 34882: contig of 704 bp in length
* 34883 34982: gap of 100 bp in length
* 34983 35682: contig of 700 bp in length
* 35683 35782: gap of 100 bp in length
* 35783 36467: contig of 685 bp in length
* 36468 36567: gap of 100 bp in length
* 36568 37251: contig of 684 bp in length
* 37252 37351: gap of 100 bp in length
* 37352 38040: contig of 689 bp in length
* 38041 38140: gap of 100 bp in length
* 38141 38838: contig of 698 bp in length
* 38839 38938: gap of 100 bp in length
* 38939 39649: contig of 711 bp in length
* 39650 39749: gap of 100 bp in length
* 39750 40426: contig of 677 bp in length
* 40427 40526: gap of 100 bp in length
* 40527 41230: contig of 704 bp in length
* 41231 41330: gap of 100 bp in length
* 41331 42022: contig of 692 bp in length
* 42023 42122: gap of 100 bp in length
* 42123 42832: contig of 710 bp in length
* 42833 42932: gap of 100 bp in length
* 42933 43636: contig of 704 bp in length
* 43637 43736: gap of 100 bp in length
* 43737 44437: contig of 701 bp in length.
Location/Qualifiers

```

```

source
1. 44437
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-38K16"
/clone_11b="RPC1-23 Female Mouse BAC"
BASE COUNT 9664 a 9556 c 9429 g 10121 t 5667 others
ORIGIN
Alignment Scores:
Pred. No.: 8.3e-24 Length: 44437
Score: 33.00 Matches: 33
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.95% Indels: 0
DB: 2 Gaps: 0
US-09-068-377c-1 (1-415) x AC100102 (1-44437)
QY 140 SerlyslsAlaTyrAspGlnLysCysArgAspAlaAspAlaGlnGlnAlaPheGlu 159
|||||
Db 44079 TCACAGAGGCGCATGTGACACAGACTGCAGGATGCAGATGTGCTGACAGCGCTTCGAG 44020
|||||
QY 160 ArgvalSerAlaasnGlyHisGlnLysGlnValGluLys 172
|||||
Db 44019 CGTGTGACTGCCAATGCCACACAGAGCACTAGAGAAAG 43981
|||||
RESULT 12
AC101133 57518 bp DNA linear HTG 23-NOV-2001
LOCUS Mus musculus clone RP23-170K14, LOW-PASS SEQUENCE SAMPLING.
AC101133
AC101133.1 GI:17059907
VERSION
KEYWORDS
HTG: HTGS. PHASE0.
SOURCE
Mus musculus.
ORGANISM
Mus musculus.
REFERENCE
1 (bases 1 to 57518)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
AUTHORS
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE
Mus musculus, clone RP23-170K14
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 57518)
Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
Anderson,S., Barin,N., Bastien,V., Boguslavsky,L., Boukhaltier,B.,
Brown,A., Camarata,J., Campiano,A., Chang,J., Chazaro,B.,
Chopel,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fairo,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
Lamaze,R., Landers,T., Lenocky,J., Levine,R., Liu,G.,
Maclean,C., MacDonald,P., Major,J., Margolis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPeeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schjorovic,N.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trillio,J., Vassiliev,B.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wymann,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center -----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

```

Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L16002
Center clone name: 170_K14

* NOTE: This record contains 73 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 710: contig of 710 bp in length
711 810: gap of 100 bp
811 1510: contig of 700 bp in length
1511 1610: gap of 100 bp
1611 2307: contig of 697 bp in length
2308 2407: gap of 100 bp
2408 3092: contig of 685 bp in length
3093 3192: gap of 100 bp
3193 3863: contig of 671 bp in length
3864 3963: gap of 100 bp
3964 4627: contig of 664 bp in length
4628 4727: gap of 100 bp
4728 5408: contig of 681 bp in length
5409 5508: gap of 100 bp
5509 6186: contig of 678 bp in length
6187 6286: gap of 100 bp
6287 6960: contig of 674 bp in length
6961 7060: gap of 100 bp
7061 7742: contig of 682 bp in length
7743 7842: gap of 100 bp
7843 8511: contig of 669 bp in length
8512 8611: gap of 100 bp
8612 9318: contig of 707 bp in length
9319 9418: gap of 100 bp
9419 10118: contig of 700 bp in length
10119 10218: gap of 100 bp
10219 10927: contig of 709 bp in length
10928 11027: gap of 100 bp
11028 11739: contig of 712 bp in length
11740 11839: gap of 100 bp
11840 12539: contig of 700 bp in length
12540 12639: gap of 100 bp
12640 13337: contig of 698 bp in length
13338 13437: gap of 100 bp
13438 14146: contig of 709 bp in length
14147 14246: gap of 100 bp
14247 14942: contig of 696 bp in length
14943 15042: gap of 100 bp
15043 15732: contig of 690 bp in length
15733 15832: gap of 100 bp
15833 16502: contig of 670 bp in length
16503 16602: gap of 100 bp
16603 17253: contig of 651 bp in length
17254 17353: gap of 100 bp
17354 18025: contig of 672 bp in length
18026 18125: gap of 100 bp
18126 18801: contig of 676 bp in length
18802 18901: gap of 100 bp
18902 19608: contig of 707 bp in length
19609 19708: gap of 100 bp
19709 20381: contig of 673 bp in length
20382 20481: gap of 100 bp
20482 21186: contig of 705 bp in length
21187 21286: gap of 100 bp
21287 22001: contig of 715 bp in length
22002 22101: gap of 100 bp

22102 22790: contig of 689 bp in length
22791 22890: gap of 100 bp
22891 23598: contig of 708 bp in length
23599 23698: gap of 100 bp
23699 24405: contig of 707 bp in length
24406 24505: gap of 100 bp
24506 25180: contig of 675 bp in length
25181 25280: gap of 100 bp
25281 25979: contig of 699 bp in length
25980 26079: gap of 100 bp
26080 26763: contig of 684 bp in length
26764 26863: gap of 100 bp
26864 27565: contig of 702 bp in length
27566 27665: gap of 100 bp
27666 28351: contig of 686 bp in length
28352 28451: gap of 100 bp
28452 29151: contig of 700 bp in length
29152 29251: gap of 100 bp
29252 29935: contig of 684 bp in length
29936 30035: gap of 100 bp
30036 30734: contig of 699 bp in length
30735 30834: gap of 100 bp
30835 31511: contig of 677 bp in length
31512 32305: contig of 694 bp in length
32306 32405: gap of 100 bp
32406 33072: contig of 667 bp in length
33073 33172: gap of 100 bp
33173 33842: contig of 670 bp in length
33843 33942: gap of 100 bp
33943 34634: contig of 692 bp in length
34635 34734: gap of 100 bp
34735 35432: contig of 698 bp in length
35433 35532: gap of 100 bp
35533 36231: contig of 699 bp in length
36232 36331: gap of 100 bp
36332 37029: contig of 698 bp in length
37030 37129: gap of 100 bp
37130 37831: contig of 702 bp in length
37832 37931: gap of 100 bp
37932 38597: contig of 666 bp in length
38598 38697: gap of 100 bp
38698 39363: contig of 666 bp in length
39364 39463: gap of 100 bp
39464 40168: contig of 705 bp in length
40169 40268: gap of 100 bp
40269 40944: contig of 676 bp in length
40945 41044: gap of 100 bp
41045 41715: contig of 671 bp in length
41716 41815: gap of 100 bp
41816 42528: contig of 713 bp in length
42529 42628: gap of 100 bp
42629 43327: contig of 699 bp in length
43328 43427: gap of 100 bp
43428 44121: contig of 694 bp in length
44122 44221: gap of 100 bp
44222 44898: contig of 677 bp in length
44899 44998: gap of 100 bp
44999 45665: contig of 667 bp in length
45666 45765: gap of 100 bp
45766 46435: contig of 670 bp in length
46436 46535: gap of 100 bp
46536 47230: contig of 695 bp in length
47231 47330: gap of 100 bp
47331 48026: contig of 696 bp in length
48027 48126: gap of 100 bp
48127 48813: contig of 687 bp in length
48814 48913: gap of 100 bp
48914 49622: contig of 709 bp in length
49623 49722: gap of 100 bp
49723 50431: contig of 709 bp in length
50432 50531: gap of 100 bp
50532 51221: contig of 690 bp in length

* 51222 51321: gap of 100 bp
* 51322 51398: contig of 677 bp in length
* 51999 52098: gap of 100 bp
* 52099 52802: contig of 704 bp in length
* 52803 52902: gap of 100 bp
* 52903 53584: contig of 682 bp in length
* 53585 53684: gap of 100 bp
* 53685 54365: contig of 681 bp in length
* 54366 54465: gap of 100 bp

Alignment Scores:

Pred. No.: 1,05e-23 Length: 57518
Score: 33.00 Matches: 33
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.95% Indels: 0
DB: 2 Gaps: 0

US-09-068-377C-1 (1-415) x AC101133 (1-57518)

QY 140 SerLysLysAlaTyrAspGlnLysCysArgAspAlaAspAlaGlnGlnAlaPheGlu 159

Db 41919 TCCAAAGAGGCATATGACCGAGAGTGCAGGATGCAGATGATCTGACAGGCCCTTCGAG 41978

QY 160 ArgValSerAlaAsnGlnLysGlnLysGlnValGluLys 172

Db 41979 CGTGTGAGTGCCTCAATGGCCACGAGAGCAGATGAGAAAG 42017

RESULT 13

AF038602 1382 bp mRNA linear PRI 02-FEB-1999
LOCUS Homo sapiens CD2 binding protein 1 short form mRNA, complete cds.
DEFINITION AF038602
ACCESSION AF038602
VERSION AF038602.1 GI:2921548
KEYWORDS
SOURCE Homo sapiens
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1382)
AUTHORS Li,J., Nishizawa,K., An,W., Hussey,R.E., Lialios,F.E.,
Sunder-Plassmann,R. and Reinherz,E.L.
TITLE A cdcl5-like adaptor protein (CD2BP1) interacts with the CD2
cytoplasmic domain and regulates CD2-triggered adhesion
JOURNAL EMBO J. 17 (24), 7320-7336 (1998)
MEDLINE 99077800
PUBMED 9857189

REFERENCE 2 (bases 1 to 1382)
AUTHORS Li,J., Nishizawa,K., An,W., Hussey,R.E., Lialios,F.E.,
Sunder-Plassmann,R. and Reinherz,E.L.
TITLE Direct Submission
JOURNAL Submitted (15-DEC-1997) Immunobiology, Dana-Farber Cancer
Institute, 44 Binney Street, Boston, MA 02115, USA
FEATURES
source location/Qualifiers
1. 1382
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="T cell"
1. 1194
/note="CD2BP1s"
/codon_start=1
/product="CD2 binding protein 1 short form"
/protein_id="A011958.1"
/db_xref="GI:2921549"

CDS

BASE COUNT 344 a 375 c 453 g 210 t

ORIGIN

Alignment Scores:
Pred. No.: 4.77e-23 Length: 1382
Score: 31.00 Matches: 60
Percent Similarity: 96.77% Conservative: 0
Best Local Similarity: 96.77% Mismatches: 1
Query Match: 7.47% Indels: 2
DB: 9 Gaps: 0

US-09-068-377C-1 (1-415) x AF038602 (1-1382)

QY 9 AspAlaPheTyrPcysArgAspPheThrAlaHisThrGlyTyrGluValLeuLeuGlnArg 28

Db 25 GATGCTTTGGTGGCGAGGACTTCAACGCCACACGGCGCTACAGGTGCTGCGACGG 84

QY 29 LeuLeuAspGlyArgLysMetCysLysAspVal-GluGluLeuLeuArgGlnArgAla 48

Db 85 CTTCGTGATGCGCAGGAAGATGTCAGAAAGACAT-CGAGAGAGCTACTGAGCGACAGCGCCA 143

QY 48 nAlaGluGluArgTyrGlyLysGluLeuValGlnIleAlaArgLysAlaGlyGlyGln 68

Db 144 GCGGAGGAGCGGTACGGAAGAGCTGTGCGATGCGACGGAAGCGAGTGGCCAGAC 203

QY 68 rGlu 69

Db 204 GGAG 207

RESULT 14

AF038603 1428 bp mRNA linear PRI 02-FEB-1999
LOCUS Homo sapiens CD2 binding protein 1 long form mRNA, complete cds.
DEFINITION AF038603
ACCESSION AF038603
VERSION AF038603.1 GI:2921550
KEYWORDS
SOURCE Homo sapiens
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1428)
AUTHORS Li,J., Nishizawa,K., An,W., Hussey,R.E., Lialios,F.E.,
Sunder-Plassmann,R. and Reinherz,E.L.
TITLE A cdcl5-like adaptor protein (CD2BP1) interacts with the CD2
cytoplasmic domain and regulates CD2-triggered adhesion
JOURNAL EMBO J. 17 (24), 7320-7336 (1998)
MEDLINE 99077800
PUBMED 9857189

REFERENCE 2 (bases 1 to 1428)
AUTHORS Li,J., Nishizawa,K., An,W., Hussey,R.E., Lialios,F.E.,
Sunder-Plassmann,R. and Reinherz,E.L.
TITLE Direct Submission
JOURNAL Submitted (15-DEC-1997) Immunobiology, Dana-Farber Cancer
Institute, 44 Binney Street, Boston, MA 02115, USA
FEATURES
source location/Qualifiers
1. 1428
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="T cell"
1. 1251
/note="CD2BP1L"
/codon_start=1
/product="CD2 binding protein 1 long form"
/protein_id="A011959.1"
/db_xref="GI:2921551"

CDS

BASE COUNT 352 a 396 c 462 g 218 t

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame-plus-p2n model

Run on: August 3, 2003, 13:48:02 ; Search time 318 Seconds
(without alignments)
2938.929 Million cell updates/sec

Title: US-09-068-377C-1

Sequence: 1 NMAOQFRDFAWCFRFTAH.....WVERNGRGVFGVSTLEKL 415

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 112599159 residues

Word size: 1

Total number of hits satisfying chosen parameters: 4368727

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame-p2n.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US09068377/rnat.28072003_091401_15908/app_query.fasta.1.583
-DB=N.Geneseq.101002 -OPW=fastap -SOFTX=olip2n.rng -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09068377.ecgn.1.1.263 @rnat.28072003_091401_15908 -NCPU=6 -ICPU=3
-NO MAP -IARGEDUTERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAEXT=60 -DELOP=6 -DELEXT=7

Database: N.Geneseq.101002.*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	415	100.0	2100	19 AAV57973	Murine tyrosine ph
2	415	100.0	2100	21 AAB64240	DNA encoding a mur
3	415	100.0	2100	21 AAA07276	Murine PST phosphat
4	1613	7.5	1613	21 AAA07277	Human PST phosphat
5	31	7.5	1803	20 AAX89760	Nucleotide sequenc
6	31	7.5	1850	21 AAZ46151	CDNA sequence enco
7	31	7.5	1858	21 AAX89759	Nucleotide sequenc
8	29	7.0	488	24 ABL87423	Human ovarian canc
9	23	5.5	531	21 AAB52643	Eosinophil activat
10	21	5.1	65	24 ABN56362	Mouse spliced tran
11	10	2.4	47	19 AAV57966	Murine PSTIP C-te
12	10	2.4	47	19 AAA07303	PCR primer for PST
13	10	2.4	48	19 AAV57984	Murine PSTIP C-te
14	10	2.4	48	19 AAV57985	Murine PSTIP C-te
15	10	2.4	48	21 AAA07301	PCR primer for PST
16	10	2.4	48	21 AAA07302	PCR primer for PST
17	10	2.4	1433	23 ABL28579	Drosophila melanog
18	10	2.4	2036	23 ABL28578	Drosophila melanog
19	9	2.2	45	19 AAV57983	Murine PSTIP C-te
20	9	2.2	45	21 AAA07300	PCR primer for PST
21	9	2.2	2551	9 AAN81551	Bio F, Bio C and B
22	8	1.9	37	19 AAV57994	Murine PSTIP PCR
23	8	1.9	37	21 AAB64248	PCR primer for PTP
24	8	1.9	37	21 AAA07285	PCR primer for PS
25	8	1.9	42	19 AAV57974	Murine PSTIP N-te
26	8	1.9	42	19 AAV57975	Murine PSTIP N-te
27	8	1.9	42	19 AAV57976	Murine PSTIP N-te
28	8	1.9	42	19 AAV57977	Murine PSTIP N-te
29	8	1.9	42	19 AAV57978	Murine PSTIP N-te
30	8	1.9	42	19 AAV57979	Murine PSTIP N-te
31	8	1.9	42	19 AAV57981	Murine PSTIP C-te
32	8	1.9	42	21 AAA07291	PCR primer for PST
33	8	1.9	42	21 AAA07292	PCR primer for PST
34	8	1.9	42	21 AAA07293	PCR primer for PST
35	8	1.9	42	21 AAA07294	PCR primer for PST
36	8	1.9	42	21 AAA07295	PCR primer for PST
37	8	1.9	42	21 AAA07296	PCR primer for PST
38	8	1.9	42	21 AAA07298	PCR primer for PST
39	8	1.9	42	21 AAA07316	Novel human diagno
40	8	1.9	367	22 AAS37421	Human OREF ORF119
41	8	1.9	421	21 AAC75564	Human OREF polyinu
42	8	1.9	421	24 ABN23123	Human eosinophil m
43	8	1.9	609	24 ABK53363	Oligonucleotide fo
44	8	1.9	743	24 ABQ44642	Oligonucleotide fo
45	8	1.9	743	24 ABQ44643	Oligonucleotide fo

ALIGNMENTS

RESULT 1
ID AAV57973 standard; cDNA; 2100 BP.

AAV57973;
24-NOV-1998 (first entry)

DE Murine tyrosine phosphorylated cleavage furrow-associated protein cDNA.

KM Mouse; tyrosine phosphorylated cleavage furrow-associated protein;

KW PSTIP; PEST family; protein tyrosine phosphatase; murine;

KM polymerisation; actin monomer; eukaryotic cell; identification;

XX antagonist; ss.
OS Mus sp.
XX
FH Key Location/Qualifiers

CDS 682..1929
/*tag= a
/product= "PSTPIP"
/note= "tyrosine phosphorylated cleavage furrow-associated protein"

XX PN W09835037-A1.
XX PD 13-AUG-1998.
XX PF 30-JAN-1998; 98WO-US01774.
XX PR 29-SEP-1997; 97US-0938829.
XX PR 07-FEB-1997; 97US-0798415.
XX PA (GETH) GENENTECH INC.
XX PI Dowbenko DJ, Lasky LA;
XX DR WPI: 1998-447234/38.
XX DR P-PSDB: AAM71595.
XX PT New PEST-type protein tyrosine phosphatase interacting polypeptide -
XX PT nucleic acids and vectors; for inducing the polymerisation of actin
XX PT monomers in eukaryotic cells and identifying antagonists
XX PS Example 9; Page 62-63; 11pp; English.
XX CC The present sequence encodes murine tyrosine phosphorylated cleavage
XX CC furrow-associated protein (PSTPIP), which is a PEST-type protein
XX CC tyrosine phosphatase (PTP)-interacting polypeptide. PSTPIP induces the
XX CC polymerisation of actin monomers in a eukaryotic cell, by introducing
XX CC a vector containing the nucleic acid sequence encoding PSTPIP into the
XX CC cell. Assays for identifying (antagonists of PSTPIP comprise contacting
XX CC PSTPIP with the agent and monitoring the ability of PSTPIP to induce
XX CC actin polymerisation.
XX CC
XX SQ Sequence 2100 BP; 548 A; 540 C; 594 G; 418 T; 0 other;

Alignment Scores:
Pred. No.: 0 Length: 2100
Score: 415.00 Matches: 415
Percent Similarity: 100.00% Conservatave: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 19 Gaps: 0

US-09-068-377c-1 (1-415) x AAV57973 (1-2100)

QY 1 MetMetAlaGlnLeuGlnPheArgAspAlaPheTrpCysArgAspPheThrAlaHisThr 20
DB 682 ATGATGCGCCACCTGCGAGTTCGAGATGCTTCTGTGCGAGGACTTACGCGCCACACA 741
QY 21 GlyTyrGlnValLeuLeuGlnArgLeuLeuAspGlyArgLysMetCysLysAspValGlu 40
DB 742 GGGTATGAGCTCTACTCTGAGAGGCTGCTGGACGGCAGAGATGTGCAAGGATGTGGAG 801
QY 41 GlnLeuLeuArgGlnArgAlaGlnAlaGlnGluArgTyrGlyLysGlnLeuValGlnIle 60
DB 802 GAGCTGCTCAGACAGAGGCGCCAGCGCGAGAGAGTACGGGAGAGAGCTGCTGCGAGATT 861
QY 61 AlaArgLysAlaGlyLysGlnThrGlnLeuMetLysSerLeuArgThrSerPheAspSerLeu 80
DB 862 GCACGGCAAGGCTGTGGCCAGACAGAGATGAATTCCTTGAGACCTCTTGACTCCCTG 921
QY 81 LysGlnGlnThrGlnAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaLeuArgGln 100
DB 922 AAGCAGCAAAACAGAGATGTGGCGACATGCACATCAGCTGCGCTGGCCCTGCGTGA 981
QY 101 GlnLeuArgSerLeuGlnGluPheArgGluArgGlnLysGlnGlnArgLysLysTyrGln 120
DB 982 GAGCTCGGAGCGCTGAGGAGTTCCTCCAGAGAGACAGAAAGACGAGGAGAAAGATATGAG 1041

QY 121 AlaIleMetAspArgValGlnLysSerLysLeuSerLeuTyrLysLysThrMetGlnSer 140
DB 1042 GCCATCATGGAACCGTGTCCAGAGAGCAAGTTGTGCTCTTACAGAAACCATGAGAGTCC 1101
QY 141 LysLysAlaTyrAspGlnLysCysArgAspAlaAspAspAlaGlnGlnAlaPheGlnArg 160
DB 1102 AAGAGGCATATGACAGCAAGAGTGCAGGATGATGATGCTGAGCAGAGCCCTTCAGAGCT 1161
QY 161 ValSerAlaAsnGlyHisGlnLysGlnValGlnLysSerGlnAsnLysAlaLysGlnGly 180
DB 1162 GTGAGTGCAGTACGCGCCAGAGAGAGTACGAGAAAGAGCCACAGAAAGCCAGAGAGTGC 1221
QY 181 LysGlnSerAlaThrGlnAlaGlnArgValTyrArgGlnAsnIleGlnGlnLeuGlnArg 200
DB 1222 AAGAGCATGACCCACAGAGGCAAGAAAGATGTCACAGCAAAATATGCAACATCGAGAGA 1281
QY 201 AlaArgThrGlnTyrGlnGlnGlnGlnHisArgThrThrCysGlnAlaPheGlnLeuGlnGlu 220
DB 1282 GCGAGAGCCAGAGTGGAGAGAGCAGCAGCAGCTACCTGTGAGGCTTCAGTTGAGAGAG 1341
QY 221 PheAspArgLeuThrIleLeuArgAsnAlaLeuTyrValHisCysAsnGlnLeuSerMet 240
DB 1342 TTGACCGGCTCACCATCTCCGCAATGCCCTGTGGTGCACGTATACAGCTCCCATG 1401
QY 241 GlnCysValLysAspAspGlnLeuTyrGlnGlnValArgLeuThrLeuGlnGlnCysAsp 260
DB 1402 CAGTGTGTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1461
QY 261 ValGlnGlyAspIleAsnGlyPheIleGlnSerLysSerThrGlyArgGlnProGlnProAla 280
DB 1462 GTGGAAGGAGCATCATGCTTCATCCAGTCCAGAGAGCAGTGGCAGAGAGCCCCAGCT 1521
QY 281 ProValProTyrGlnAsnTyrArgAspArgGlnValThrProLeuIleGlySerProSer 300
DB 1522 CCGTCCCTTATCAGACACTATGATGACAGGAGGTGACCCCATGTTGGACGCCCTAGC 1581
QY 301 IleGlnProSerCysGlyValIleLysArgPheSerGlyLeuLeuHisGlySerProLys 320
DB 1582 ATCCAGCCCTCCGCGTGTGATAAAGAGTCTCTGCGCTCTACATGGAAGTCCCAAG 1641
QY 321 ThrThrProSerAlaProAlaIleSerThrGlnThrLeuThrProThrProGlnArgAsn 340
DB 1642 ACCACACCTTCTGCTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1701
QY 341 GlnLeuValTyrAlaSerIleGlnValGlnAlaThrGlnGlnLysAsnLeuAsnSerSerAla 360
DB 1702 GAGTGTGCTACGATCCATGCAAGTGCAGGCGACCCAGAGAAACCTTAACATCATCAGCC 1761
QY 361 GlnAspTyrArgAlaLeuTyrAspTyrThrAlaGlnAsnSerAspGlnLeuAspIleSer 380
DB 1762 CAGGACTCCGGGCACTTACGACTACACTGACAGAAATTCGATGAGCTGCACATTTCC 1821
QY 381 AlaGlyAspIleLeuAlaValIleLeuGlnGlnGlnLysAspGlyTyrThrThrValGlnArg 400
DB 1822 GCGGAGACATCTCTGCGGCTCATCTCGAAGGAGAGATGGCTGTGGACTGTGGAGCGG 1881
QY 401 AsnGlyGlnArgGlyPheValProGlySerTyrLeuGlnLysLeu 415
DB 1882 AACGAGCAACGCTGTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1926

RESULT 2
AAA64240 standard; DNA; 2100 BP.
AAA64240:
XX 20-DEC-2000 (first entry)
XX DNA encoding a murine PTP phosphatase interacting protein (PSTPIP).
XX protein tyrosine phosphatase; PTP; phosphatase interacting protein;
XX PSTPIP; PEST family; protein tyrosine phosphatase; actin monomer;
XX tissue typing; tumour cell; tumour imaging; ss.

XX Mus sp.
XX Key Location/Qualifiers
FH CDS 682..1929
FT /tag- a
FT /product- "protein tyrosine phosphatase (PTP)
FT phosphatase interacting protein (PSTPIP)"
XX
XX US611073-A.
XX
XX 29-AUG-2000.
XX
XX 06-FEB-1998: 98US-0020222.
XX
XX 07-FEB-1997: 97US-0104590.
XX
XX (GENTH) GENENTECH INC.
XX
XX Lasky LA;
XX
XX WPI: 2000-586378/55.
XX P-PSDB: AAB08486.
XX
XX Novel PTP phosphatase interacting protein useful for inducing
XX polymerisation of actin monomers and for identifying homologue of PTP
XX phosphatase interacting protein
XX
XX Disclosure: Column 47-50; 48pp; English.
XX
XX The present sequence encodes a protein tyrosine phosphatase (PTP)
XX phosphatase interacting protein (PSTPIP). PSTPIP polypeptides are
XX bound by and dephosphorylated by the PEST family of protein tyrosine
XX phosphatases. PSTPIP associates with actin. PSTPIP is useful for
XX inducing the polymerisation of actin monomer in eukaryotic cells by
XX introducing the polypeptide into the cell. The polypeptide is useful
XX for identifying and isolating PSTPIP homologues in another mammalian
XX species, in screening assays to identify antagonists and agonists of
XX native PSTPIP polypeptide and as molecular weight markers on protein
XX gels. The PSTPIP nucleic acid is useful for tissue typing of specific
XX mammalian tissues, for preparing PSTPIP polypeptides by recombinant
XX techniques, as hybridisation probes for searching cDNA and genomic
XX libraries for the coding sequence of other PSTPIP analogues and to
XX isolate homologous genes specifically expressed in tumour cells.
XX Antagonists of PSTPIP peptide are useful for inhibiting biological
XX activity of the peptide. Antibodies of PSTPIP are useful to identify
XX rapidly dividing cells and are used to image tumours comprising such
XX rapidly dividing cells.
XX
XX SQ Sequence 2100 BP; 548 A; 540 C; 594 G; 418 T; 0 other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 0 Length: 2100
XX Score: 415.00 Matches: 415
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 100.00% Indels: 0
XX DB: 21 Gaps: 0
XX
XX US-09-068-377C-1 (1-415) x AAA64240 (1-2100)
XX
XX QY 1 MetMetAlaGlnLeuGlnPheArGAsPaAlaPheTrpCysArGAsPheThrAlaHisThr 20
XX |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
XX DB 682 ATGATGGCCAGCTCGAGTCCGAGATGCTTCTGTGTCAGGACATTCACGCCACACA 741
XX
XX QY 21 GlyTyrGlnValLeuLeuGlnLnaArgLeuLeuAspGlyArgLysMetCysLysAspValGlu 40
XX |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
XX DB 742 GGGTATGAGGTCTACTCGAGAGGCTGTCGACGCCAGCAAGATGTCGAAGCATGTGAG 801
XX
XX QY 41 GluLeuLeuArgGlnArgLnaGlnAlaGlnGluArgTyrGlyLysLnuLeuValGlnIle 60
XX |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
XX DB 802 GAGCTGCTCAGACAGAGGCCCGACGGCGAGAGAGGTACGGGCAAGAGAGTGTGACAGATT 861

QY 61 AlaArgLysAlaGlyGlnThrGluMetAsnSerLeuArgThrSerPheAspSerLeu 80
|||
DB 862 GCACGCAGAGCTGTGGCCACAGACAGATGAATTCCTTGAGACCTCTTGTACTCCCTG 921
QY 81 LysGlnGlnThrGluAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaLeuArgGlu 100
|||
DB 922 AACGACGAACAGAGATGTGGCGACATGCACATCCAGCTGGCCCTGGCCGCTGAG 981
QY 101 GluLeuArgSerLeuGlnGluPheArgGluArgGlnLysGlnAlaArgLysLysTyrGlu 120
|||
DB 982 GAGCTGCGGAGCCTGGAGGAGTTCCGAGACAGACAGAAAGACGCGAAGCATGTAG 1041
QY 121 AlaIleMetAspArgValGlnLysSerLysLeuSerLeuTyrLysThrMetGluSer 140
|||
DB 1042 GCCATCATGAGCCGTGTCCAGAAAGACAACTTGTCTCTACAGAAAGACCATGAGACTCC 1101
QY 141 LysLysAlaTyrAspClnLysCysArgAspAlaAspAspAlaGlnGlnAlaPheGluArg 160
|||
DB 1102 AAGAGGCATATGACACAGAGAGTCAGAGATGATGATGATGATGATGATGATGATGATGAT 1161
QY 161 ValSerAlaAsnGlnLysGlnLysGlnValGlnLysSerGlnAsnLysAlaLysGlnCys 180
|||
DB 1162 GTGAGTGCATATGACACAGAGAGTCAGAGATGATGATGATGATGATGATGATGATGATGAT 1221
QY 181 LysGlnSerAlaThrGlnAlaGluArgValTyrArgGlnAsnIleGlnGlnLeuGluArg 200
|||
DB 1222 AAGAGTGCATATGACACAGAGAGTCAGAGATGATGATGATGATGATGATGATGATGATGAT 1281
QY 201 AlaArgThrGluTrpGlnGlnGlnLysArgThrThrCysGlnAlaPheGlnLeuGlnGlu 220
|||
DB 1282 GCGAGGACCGAGTGGGAGCGAGAGACCGGAGTACCTGTGAGGCTTCGAGTGGAGAGAG 1341
QY 221 PheAspArgLeuThrIleLeuArgAsnAlaLeuThrValHisCysAsnGlnLeuSerMet 240
|||
DB 1342 TTGGACCGGTCTCCATCCCTCCCAATGCCCTGTGGTGGACGTAAACAGCTCTCATG 1401
QY 241 GlnCysValLysAspAspGlnLeuTyrGlnGluValArgLeuThrLeuGlnGlnCysAsp 260
|||
DB 1402 CACTGTGTCTCAGATATATGAGCTCTATGAGAGAAAGTGGCTGACCTTGAGGGCTGTGAT 1461
QY 261 ValGlnGlyAspIleAsnGlyPheIleGlnSerLysSerThrGlyArgGluProAla 280
|||
DB 1462 GTGGAAGGTGACATCAATGAGCTTCATCCATCCAGAGAGAGAGAGAGAGAGAGAGAGAG 1521
QY 281 ProValProGlnGlnAsnTyrTyrAspArgGluValThrProLeuIleGlySerProSer 300
|||
DB 1522 CCGGTGCTTATACAGACTACTATGACAGGAGGTGACCCCATGATTGGCAGCCCTTAC 1581
QY 301 IleGlnProSerCysGlyValIleLysArgPheSerCysLeuLeuHisGlySerProLys 320
|||
DB 1582 ATCCAGCCCTCTCGCGGTGTGATTAAGAGGTCTCTGTGGCTGTACATGGAATGCCAAG 1641
QY 321 ThrThrProSerAlaProAlaAlaSerThrGluThrLeuThrProThrProGluArgAsn 340
|||
DB 1642 ACCACACCTTCT 1701
QY 341 GluLeuValTyrLysSerIleGlnValGlnAlaThrGlnLysLnuLeuAsnSerSerAla 360
|||
DB 1702 GAGTGTGCTTACCATTCATCGAAGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1761
QY 361 GlnAspTyrArgAlaLeuTyrArgAspTyrThrAlaGlnAsnSerAspGluLeuAspIleSer 380
|||
DB 1762 CAGAGACACCGGGACACTCTACGACTACACTACGACAGATTCGATGAGCTGGACATTTCC 1821
QY 381 AlaGlyAspIleLeuAlaValIleLeuGlnGlyLnuAspGlyTyrTrpThrValGluArg 400
|||
DB 1822 GCGGAGACATTCCTGGGGGTCAATCTGGAAGGGAGAGATGGCTGTGAGTGTGAGAGCGG 1881
QY 401 AsnGlyLnaArgLysPheValProGlySerTyrLeuGlnLysLeu 415
|||
DB 1882 AACGACAAAGCTGTGTCTTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1926

RESULT 3

```

AAA07276      ID    AAA07276 standard; DNA; 2100 BP.
XX AC          AAA07276;
XX XX          27-JUN-2000 (first entry)
XX DT
XX DE           Murine PST phosphatase interacting protein coding sequence.
XX RN
XX RS           PST phosphatase interacting protein; PSTRIP; tumour therapy;
XX RW           protein tyrosine phosphatase; mouse; ss.
XX OS           Mus sp.
XX PN           US6040437-A.
XX PD           21-MAR-2000.
XX PF           29-SEP-1997;   97US-0938830.
XX PR           07-FEB-1997;   97US-0104590.
XX RA           (GETH ) GENENTECH INC.
XX PI           Dowbenko DJ , Lasky LA;
DR MPD:         2000-282393/24.
PS             P-PSTDB; AAV81928.
PT PT           Novel genes encoding protein tyrosine phosphatase binding proteins
PR provide more specific targets for tumour therapy -
PS PS           Claim 5; Column 63-66; 65pp; English.
CC CC           This sequence represents the pstr phosphatase interacting protein (PSTRIP)
CC coding sequence of the invention. The protein is a protein tyrosine
CC phosphatase that possesses a non-catalytic domain comprising a proline,
CC serine and threonine rich region and a C-terminal segment of 20 amino
acid (aa's) rich in proline, and defines an SH3 binding domain. Nucleic
acids encoding native PSTRIP molecules can be used to isolate homologous
genes specifically expressed in tumour cells, which might provide more
specific targets for tumour therapy. The DNA is also useful for the
preparation of PSTRIP polypeptides by recombinant techniques and as
hybridisation probes for searching cDNA and genomic libraries for the
coding sequence of other PSTRIP polypeptide analogues in other species.
SQ SQ           Sequence 2100 BP; 548 A; 540 C; 594 G; 418 T; 0 other;

Alignment Scores:
Pred. No.:       0               Length:     2100
Score:           415.00        Matches:      415
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches:    0
Query Match:     100.00%      Indels:      0
DB:              21            Gaps:         0

US-09-068-377C-1 (1-415) x AAA07276 (1-2100)
QY 1 MetMetAlaGlnLeuGlInpheArgAspaIaphETPCysArGaSpheThrAlaHisThr 20
Db | |||||
682 ATGATGGGCCACAGCTGCAGTTCGAGATGCCCTTGCGTAGGGAAGACTTCACGGCCCACA 741
QY 21 GlyTYRGIValleuleuLenglArGTdeuleulaSPGLYArdgylsmetCySLYSASpVaIGlu 40
Db | |||||
742 GGGRATGAGGTCTCACTCGAGAAGCGCTGTGGAGCGGAGCAAAGATGTGCCAAGATTGTGGAG 801
QY 41 GlutleuluArgGlnArgAlaglnalaglUGlAgRTyrGISlysgIUeuaValglNile 60
Db | |||||
802 GAGCTGCTTAAGACAGAGGGCCCCGAGGAGAAAGTTACGGGAAGAAGCTGTGTCCAATT 861
QY 61 AlaArgVlaIaGlylgIntnrgIUmetasnSerleuarqTrtSerpeaspSsrLeu 80

```

Db	862	GCACGCCAAGCGTGGTGGCCACAGACAGATGCAATTCCTCGAGACACTCTTGATCCCTG	921
Qy	81	LysGlnGlnThrGluAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaGlu	100
Db	922	AAGCAGCAAAACAGAGAAATGTCGGCCAGTGCACACATCCAGCTGGCGCCCTGCGCTGGCAG	981
Qy	101	GluLeuArgSerLeuGluGluPheArgGluArgGlnGlnLysGlnAlaGlyLysTyrGlu	120
Db	982	GAGCTGGCGAGCTCGAGAGGAGTTCCGAGAGAGACAGAAAGACGACCGGACAGACTATGAC	1041
Qy	121	AlaIleMetAspArgValGlnLysSerLysLeuSerLeuTyrLysTyrMetGluSer	140
Db	1042	GCCATCATGAGACCGTGTCCAGAAAGACAGCAAGTTGCTGCTCTACAGAAAGACCATTGAGATGC	1101
Qy	141	LysLysAlaTyrAspGlnLysCysArgAspAlaAspAspAlaGluGlnAlaPheGluArg	160
Db	1102	AAGAAGGCAATGTGACCCAGAAAGTGCAGAGATGCAATATGCTGTGACAGGCTTGACGCT	1161
Qy	161	ValSerAlaAsnGlyHisGlnLysGlnValGlnLysSerGlnAsnLysAlaLysGlnCys	180
Db	1162	GTGAGTGGCATGGCCACACCAACAGCAAGTACAAAAGAGCCAGAACCAAGCCACAGCTGC	1221
Qy	181	LysGlnSerAlaThrGluAlaGluArgValTyrArgGlnAsnIleGluGlnLeuGluArg	200
Db	1222	AAGGAGTCAGCCACAGAGGACAGAAAGACTGTACAGGCAAAATATCGAAACCACTGGAGAGA	1281
Qy	201	AlaArgThrGluTyrPgluGlnGluHisArgThrLysCysGluAlaPheGlnLeuGlnGlu	220
Db	1282	CGCAGAGCCAGTGGGAGCAGAGACCCGCGACTACTGTGAGGCTTCGCACTTCAGAG	1341
Qy	221	PheAspArgLeuThrIleLeuArgAsnAlaLeuTyrValHisCysAsnGlnLeuSerMet	240
Db	1342	TTTGACCGCGCTCACCATCTCCGCAATGCCCTGTGGGTGCATCTATACACACTCTCCATG	1401
Qy	241	GlnCysValLysAspAspGluLeuTyrGlnGluValArgLeuThrLeuGlnGlyCysAsp	260
Db	1402	CAGTGTGTCAGAGATGATGACTCTATGAGAAAGTGGCGCTGACCTTGAGAGCGTGTAT	1461
Qy	261	ValGlnGlyAspIleAsnGlyPheIleGlnSerLysSerThrGlyArgGluProProAla	280
Db	1462	GTGGAGAGTGCATCAATCAATGCGCTTATCCAGTCCAAAGACACTGGCAGAGACCCCGAGCT	1521
Qy	281	ProValProTyrGlnAsnTyrTyrAspArgGluValThrProLeuIleGlySerProSer	300
Db	1522	CCGGGCTTATCCAGACTACTATCAGGAGGAGGACCCACGATGATGGACGCCCTATGC	1581
Qy	301	IleGlnProSerCysGlyValIleLysArgPheSerGlyLeuLeuHisGlySerProLys	320
Db	1582	ATCCAGCCCTCCTCGGTGGTGTAAAGAGTTCTCTGGCTGCTACTACATGAGAATGCCAAG	1641
Qy	321	ThrThrProSerAlaProAlaAlaSerThrGluThrLeuThrProThrProGluArgAsn	340
Db	1642	ACCACACTTCTGCTCTCGTGGCTTCCACAGAGACTGTGACTCCACCCCTGAGGCGCAT	1701
Qy	341	GluLeuValTyrAlaSerIleGluValGlnAlaThrGlnGlyAsnLeuAsnSerSerAla	360
Db	1702	GAGTTGGGTCAAGCATTCATGAGAGTGCAGGCGACCCAGGGAACCTTAACATCATCAGCC	1761
Qy	361	GlnAspTyrArgAlaLeuTyrAspTyrThrAlaGlnAsnSerAspGluLeuAspIleSer	380
Db	1762	CAGGACTACCGGGCACTCTACGACTACACTCCACAGAAATTCTGATGACTGGACATTTCC	1821
Qy	381	AlaGlyAspIleLeuAlaValIleLeuGlnGlyGluAspGlyTyrPTPTThrValGluArg	400
Db	1822	GCGGAGACATCTGGCGGTATCTCGAAGAGGAGCATGGCTGGTGACTCTGTGAGCGG	1881
Qy	401	AsnGlyGlnArgGlyPheValProGlySerTyrLeuGlnLysLeu	415
Db	1882	AACGACACAACGTGGCTTGTCTCCTGGTGCATCTGAGAGAGCTC	1926

```

XX AAA07277;
AC
XX
XX 27-JUN-2000 (first entry)
DE Human PST phosphatase interacting protein coding sequence.
XX
XX PST phosphatase interacting protein; PSTPIP; tumour therapy;
KW protein tyrosine phosphatase; human; ss.
XX
OS Homo sapiens.
XX
XX US6040437-A.
XX
XX 21-MAR-2000.
XX
XX 29-SEP-1997; 97US-0938830.
XX
XX 07-FEB-1997; 97US-0104590.
XX
XX (GETH ) GENENTECH INC.
XX
XX Dowdenko DJ, Lasky LA;
XX
XX WPI: 2000-282393/24.
XX
XX P-PSDB; AAY81929.
XX
XX
XX Novel genes encoding protein tyrosine phosphatase binding proteins
PT useful for isolating homologous genes, e.g. in tumour cells, which
PT provide more specific targets for tumour therapy
XX
XX
XX Claim 5; Column 79-82; 65pp; English.
XX
XX This sequence represents the PST phosphatase interacting protein (PSTPIP)
CC coding sequence of the invention. The protein is a protein tyrosine
CC phosphatase that possesses a non-catalytic domain comprising a proline,
CC serine and threonine rich region and a C-terminal segment of 20 amino
CC acid (aa's) rich in proline, and defines an SH3 binding domain. Nucleic
CC acids encoding native PSTPIP molecules can be used to isolate homologous
CC genes specifically expressed in tumour cells, which might provide more
CC specific targets for tumour therapy. The DNA is also useful for the
CC preparation of PSTPIP polypeptides by recombinant techniques and as
CC hybridisation probes for searching cDNA and genomic libraries for the
CC coding sequence of other PSTPIP polypeptide analogues in other species.
XX
XX
SQ Sequence 1613 BP; 361 A; 452 C; 535 G; 265 T; 0 other:
SQ
Alignment Scores:
Pred. No.: 1.44e-21 Length: 1613
Score: 31.00 Matches: 60
Percent Similarity: 96.77% Conservative: 0
Best Local Similarity: 96.77% Mismatches: 1
Query Match: 7.47% Indels: 2
DB: Gaps: 0
21
US-09-068-377c-1 (1-415) x AAA07277 (1-1613)
OY 9 AspalaphetPCysArgAspPheThrAlaHisThrGlyTyrGluValLeuLeuGlnArg 28
Db 368 GATCCTTTTGCTGCTCAGGACTTCACACCCACAGGGCTACGAGGCTCTCTGAGGGG 427
OY 29 LeuLeuAspGlyArgLysMetCysLysAspVal-GluGluLeuLeuArgGlnArgAlaG1 48
Db 428 CTTCTGATGACGACGAGAGATGTCGAAGACAT-AGAGAGAGCTACTGAGGACGAGGGCCCA 486
OY 48 nalaagluGluArgTyrGlyLysGluLeuValGlnIleAlaArgLysAlaGlyGlnThr 68
Db 487 GGGCGAGAGCGGTACGGGAGAGAGCTGTGTCAGATCGCACGGAAGCGAGGTGGCCAGAC 546
OY 68 rGlu 69
Db 547 GGAG 550

```

```

RESULT 5
AAx89760
ID AAX89760 standard; DNA; 1803 BP.
XX
XX
XX AAX89760;
AC
XX
XX 28-SEP-1999 (first entry)
DE
XX
XX Nucleotide sequence of CD2BP1S.
DE
XX
XX nucleotide; DNA; CDS; CD2; immune response; lymphocytes;
KW variant; cell adhesion; ds.
XX
XX
XX Homo sapiens.
OS
XX
XX
XX Key Location/Qualifiers
XX CDS 440..1630
XX FT /*tag= "a
XX FT /product= "CD2BP1S"
XX FT /note= "variant of CD2BP1"
XX
XX
XX WO9936534-A1.
XX
XX
XX 22-JUL-1999.
XX
XX
XX 14-DEC-1998; 98WO-US26699.
XX
XX
XX 13-JAN-1998; 98US-0006428.
XX
XX
XX (DAND ) DANA FARBER CANCER INST INC.
XX
XX An W, Li J, Nishizawa K, Reinherz EL;
XX
XX WPI: 1999-444396/37.
XX
XX
XX A human CD2 cytoplasmic tail binding protein, CD2BP1 and related
PT polynucleotides, useful in identifying agents which modulate signal
PT transduction, cell adhesion or motility
XX
XX
XX Claim 11; Page 102-104; 107pp; English.
XX
XX
XX This is the nucleotide sequence for a variant of CD2BP1, a cd215-like
CC adapter protein. Anti-CD2BP1 antibodies can be used to assay for the
CC presence of CD2BP1 in a cell or tissue sample.
CC
CC Modulatory agents identified in the methods are used to inhibit or
CC enhance CD2BP1 activity. These agents can be used to modulate
CC signal transduction, cell adhesion or motility. CD2BP1 inhibitors are
CC useful for enhancing signal transduction, cell adhesion or
CC motility. CD2BP1 enhancers have the opposite effect. Mimics of CD2BP1
CC protein activity can also be used to modulate signal transduction, etc
CC (all claimed).
CC
CC Inhibition of CD2BP1 is useful for cancer therapy, to augment the
CC immune response to cancer cells and tumors. Enhancement of CD2BP1 is
CC useful in immunocompromised individuals and in immunodeficiency related
CC diseases. Such diseases and conditions, include rheumatoid arthritis,
CC juvenile diabetes, systemic lupus erythematosus, and transplantation, to
CC reduce adhesion of the recipient's T cells with tissues in the
CC autograft.
XX
XX
SQ Sequence 1803 BP; 381 A; 540 C; 585 G; 297 T; 0 other:
SQ
Alignment Scores:
Pred. No.: 1.6e-21 Length: 1803
Score: 31.00 Matches: 60
Percent Similarity: 96.77% Conservative: 0
Best Local Similarity: 96.77% Mismatches: 1
Query Match: 7.47% Indels: 2
DB: Gaps: 0
20
US-09-068-377c-1 (1-415) x AAX89760 (1-1803)
OY 9 AspalaphetPCysArgAspPheThrAlaHisThrGlyTyrGluValLeuLeuGlnArg 28
Db

```

Db 464 GATGCTTTTGGTCAGAGGACTTCACAGCCACAGGGGCTGACAGGTCTGTCAGCGG 523
QY 29 LeuLeuAspGlyArgIysMetCysLysAspVal-GluGluLeuLeuArgGlnArgAlaG1 48
Db 524 CTTCGTGATGACGAGGAAGATGTGCAGAAAGACAT-GGAGGAGCTACTGAGCGAGAGGCCCA 582
QY 48 nalaGluGluArgTyrGlyIysGluLeuValGlnIleAlaArgLysAlaGlyGlyGlnH 68
Db 583 GGGCGAGAGAGCGGTACGGAGGAGCTGTGTGACAGATCGACGGAGGCGAGGTGCCAGAC 642
QY 68 rGlu 69
Db 643 GGAG 646
RESULT 6
AAZ46151
ID AAZ46151 standard; cDNA: 1850 BP.
XX
AC AAZ46151;
DT 16-MAY-2000 (first entry)
XX
DE cDNA sequence encoding a human phosphorylation effector PHSP-14.
XX
KM Human: phosphorylation effector; PHSP; proliferative disorder;
KM immune disorder; neuronal disorder; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 431..1681
FT /*tag= a
FT /product= "phosphorylation effector"
FT
XX W0200006728-A2.
XX
PD 10-FEB-2000.
XX
PE 28-JUL-1999; 99WO-US17132.
XX
PR 28-JUL-1998; 98US-0123494.
PR 14-SEP-1998; 98US-0152814.
PR 14-OCT-1998; 98US-0173482.
PR 03-NOV-1998; 98US-0106889.
PR 19-NOV-1998; 98US-0109093.
PR 22-DEC-1998; 98US-0113796.
PR 12-JAN-1999; 99US-0173482.
PR 12-JAN-1999; 99US-0229005.
XX
XX (INCY-) INCYTE PHARM INC.
XX
PI Hillman JL, Lal P, Tang YT, Corley NC, Guegler KJ, Baughn MR;
PI Patterson C, Bandman O, Au-Young J, Gorgone GA, Yue H, Azimzal Y;
PI Reddy R, Lu DM, Shih LL;
XX
DR WPI: 2000-183125/16.
DR P-PSDB; AAY68782.
XX
PT New human phosphorylation effectors useful for the diagnosis, treatment
PT and prevention of proliferative, immune and neuronal disorders
XX
PS Claim 9; Page 128-129; 142pp; English.
XX
CC AAZ46138-246168 encode human phosphorylation effectors (PHSP),
CC designated PHSP1-PHSP3 (the protein sequence for PHSP28 is not given
CC in the specification). The sequences were isolated from cDNA libraries
CC prepared from various human tissues. The PHSP proteins are useful for
CC the diagnosis, treatment and prevention of proliferative disorders,
CC immune disorders and neuronal disorders. The PHSP proteins form
CC pharmaceutical compositions which useful for treating or preventing
CC disorders associated with decreased PHSP expression/activity. PHSP
CC antagonists are useful for treating or preventing disorders associated
CC with increased PHSP expression/activity.

XX
SQ Sequence 1850 BP; 395 A; 559 C; 590 G; 306 T; 0 other;
Alignment Scores:
Pred. No.: 1.63e-21 Length: 1850
Score: 31.00 Matches: 31
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.47% Indels: 0
DB: 21 Gaps: 0
US-09-068-377c-1 (1-415) x AAZ46151 (1-1850)
QY 204 GluTrpGluGlnGluHisArgThrCysGluAlaPheGlnLeuGlnGluPheAspArg 223
Db 1040 GAGTGGAGCAGAGAGCAGCCGAGCACCCTGTGAGGCTTTAGCTGCAAGAGTTGACCGG 1099
QY 224 LeuThrIleLeuArgAsnAlaLeuTrpValHis 234
Db 1100 CTGACCATCTCCGCAACGCCCTGTGGTGCAC 1132
RESULT 7
AAx89759
ID AAX89759 standard; DNA: 1858 BP.
XX
AC AAX89759;
XX
DT 28-SEP-1999 (first entry)
XX
DE Nucleotide sequence of CD2BP1L.
XX
DE nucleotide; DNA; CDS; CD2; immune response; lymphocytes;
XX
KW variant; cell adhesion; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 440..1687
FT /*tag= a
FT /product= "CD2BP1L"
FT /note= "Variant of CD2BP1"
XX
XX W09936534-A1.
XX
PD 22-JUL-1999.
XX
PE 14-DEC-1998; 98WO-US26699.
XX
PR 13-JAN-1998; 98US-0006428.
XX
XX (DAND) DANA FARBER CANCER INST INC.
XX
PI An W, Li J, Nishizawa K, Reinherz EL;
XX
DR WPI: 1999-444396/37.
XX
PT A human CD2 cytoplasmic tail binding protein, CD2BP1 and related
PT polynucleotides, useful in identifying agents which modulate signal
PT transduction, cell adhesion or motility
XX
PS Claim 9; Page 98-100; 107pp; English.
XX
CC This is the nucleotide sequence for a variant of CD2BP1, a cdc15-like
CC adapter protein. Anti-CD2BP1 antibodies can be used to assay for the
CC presence of CD2BP1 in a cell or tissue sample.
CC Modulatory agents identified in the methods are used to inhibit or
CC enhance CD2BP1 activity. These agents can be used to modulate
CC signal transduction, cell adhesion or motility. CD2BP1 inhibitors are
CC useful for enhancing signal transduction, cell adhesion or
CC motility. CD2BP1 enhancers have the opposite effect. Mimics of CD2BP1
CC protein activity can also be used to modulate signal transduction, etc
CC (all claimed).
CC Inhibition of CD2BP1 is useful for cancer therapy, to augment the

CC Immune response to cancer cells and tumors. Enhancement of CD2Bp1 is
CC useful in immunocompromised individuals and in immunodeficiency related
CC diseases. Such diseases and conditions, include rheumatoid arthritis,
CC juvenile diabetes, systemic lupus erythematosus, and transplantation, to
CC reduce adhesion of the recipient's T cells with tissues in the
CC autograft.

XX Sequence 1858 BP; 398 A; 561 C; 594 G; 305 T; 0 other;

Alignment Scores:

Pred. No.: 1,64e-21 Length: 1858
Score: 31.00 Matches: 60
Percent Similarity: 96.77% Conservative: 0
Best Local Similarity: 96.77% Mismatches: 1
Query Match: 7.47% Indels: 2
DB: Gaps: 0

US-09-068-377c-1 (1-415) x AAX89759 (1-1858)

OY 9 AspalahetpCYsArGAspPheThraLahHstHrGLyTYrGLuValLeuLeuGlnArg 28
DB 464 GATGCTTTGGTGCAGGACTTCACGCCACAGGGCTAGAGGTGCTGTCAGCGG 523
OY 29 LeuLeuAspGLyArGLyMetCylsAspVal -GluGluLeuLeuArgGlnArgAlaG1 48
DB 524 CTCTGATGCGAGGAGATGTGCMAAGACAT -GGAGAGACTACTGAGGACAGGCCCA 582
OY 48 nAlaGluGluArgTYrGLyGLyGluLeuValGlnLeuAlaArgLysAlaGlyGlnTh 68
DB 583 GCGGAGAGCGGTACGGGAGGAGACTGTGCGATGCGACGAGGAGGAGGTGGCCAGAC 642
OY 68 rGlu 69
DB 643 GGAG 646

RESULT 8

ABL87423/c
ID ABL87423 standard; cDNA: 488 BP.

XX- ABL87423;

DT 17-MAY-2002 (first entry)

XX Human ovarian cancer related cDNA clone SEQ ID NO:10401.

XX Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.

XX Homo sapiens.

XX WO200192581-A2.

XX 06-DEC-2001.

XX 29-MAY-2001; 2001WO-US17756.

XX 26-MAY-2000; 2000US-207484P.

XX (CORI-) CORIXA CORP.

XX Algate PA, Harlocker SL, Jones R;

XX WPI; 2002-122075/16.

XX Composition for therapy and diagnosis of ovarian cancer comprising
XX polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
XX polypeptide, antibody specific to polypeptide or T cell expressing
XX polypeptide
XX Claim 1; SEQ ID 10401; 489pp; English.

XX The present invention describes a composition (I) comprising: carriers
XX and immunostimulants; and a polypeptide (II) of a ovarian tumour
XX polypeptide encoded by a polynucleotide (III) having a cDNA sequence

CC (S1) from the 10912 nucleotide sequences as given in ABL77023 to
CC ABL87934, (III) encoding (II) having a sequence (S2), a T cell
CC population of (II), or antigen presenting cells that express (II).
CC (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to
CC (S1) can be used for detecting ovarian cancer in a patient's biological
CC sample preferably serum or ovarian tissue. The method comprises
CC contacting a biological sample from a patient with (IV), detecting the
CC amount of polynucleotide hybridising to (IV) and comparing the amount to
CC a predetermined cutoff value and thereby detecting ovarian cancer in the
CC patient, where the amount of polynucleotide hybridising to (IV) is
CC detected preferably by polymerase chain reaction (PCR). (I) comprising
CC (III) and/or (II) is useful for stimulating and/or expanding T cells
CC specific for an ovarian tumour protein comprising contacting T cells
CC with (III) or (II). (III) is useful in design and preparation of
CC ribozyme molecules for inhibiting expression of the tumour polypeptides
CC and proteins in tumour cells; and to isolate a full length gene from a
CC suitable library e.g., a tumour cDNA library using well known
CC techniques.

XX Sequence 488 BP; 77 A; 147 C; 139 G; 123 T; 2 other;

Alignment Scores:

Pred. No.: 5,07e-20 Length: 488
Score: 29.00 Matches: 29
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.99% Indels: 0
DB: Gaps: 0

US-09-068-377c-1 (1-415) x ABL87423 (1-488)

OY 387 ValIleLeuGluGLyGluAspGLyTYrTPThraValGluArgAsnGLyLnaArgGlyPhe 406
DB 270 GTGATCTGGAAGGAGGATGCTGTGACTGTGAGAGGAGGAGCGAGGCTGCTTC 211
OY 407 ValProGlySerTYrLeuGluLysLeu 415
DB 210 GTCCCTGGTTCCTACTGTGAGAAAGCTT 184

RESULT 9

AAA52643
ID AAA52643 standard; DNA: 531 BP.

XX- AAA52643;

DT 07-DEC-2000 (first entry)

XX Eosinophil activation cell adhesion peptide gene 1.

XX Eosinophil activation; human; allergy; eosinophils; cancer;

XX inflammation; signalling peptide; cell adhesion peptide;

XX G-protein coupled receptor; secreted protein; cell surface protein; ds.

XX Homo sapiens.

XX WO200032630-A2.

XX 08-JUN-2000.

XX 06-DEC-1999; 99WO-US28773.

XX 04-DEC-1998; 98US-0111006.

XX (SEAR) SEARLE & CO G D.

XX Dotson SB, Ma X;

XX WPI; 2000-465041/40.

XX Novel nucleic acids derived from activated eosinophil cells useful for
XX treating allergic diseases such as asthma comprises a specific
XX nucleotide sequence

DR WPI; 1998-447234/38.
XX
XX New PEST-type protein tyrosine phosphatase interacting polypeptide -
PT nucleic acids and vectors; for inducing the polymerisation of actin
PT monomers in eukaryotic cells and identifying antagonists
XX
XX Example 6; Page 44; 11pp; English.
XX
CC The present sequence represents a PCR primer involved in deletion
CC mutagenesis of murine tyrosine phosphorylated cleavage furrow-associated
CC protein (PSTPIP), which is a PEST-type protein tyrosine phosphatase
CC (PTP)-interacting polypeptide. PSTPIP induces the polymerisation of
CC actin monomers in a eukaryotic cell, by introducing a vector containing
CC the nucleic acid sequence encoding PSTPIP into the cell. Assays for
CC identifying (antagonists of PSTPIP comprise contacting PSTPIP with the
CC agent and monitoring the ability of PSTPIP to induce actin
CC polymerisation.
XX
XX SQ Sequence 47 BP; 9 A; 14 C; 15 G; 9 T; 0 other;
Alignment Scores:
Pred. No.: 0.135 Length: 47
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.41% Indels: 0
DB: 19 Gaps: 0
US-09-068-377C-1 (1-415) x AAV57986 (1-47)
QY 91 HisIIegInLeuAlaLeuAlaLeuArgGlu 100
DB 47 CACATCCAGCTGGCCCTGGCCTGCTAG 18
RESULT 12
AA07303/C
ID AAA07303 standard; DNA; 47 BP.
XX
XX AAA07303;
AC
XX
XX 27-JUN-2000 (first entry)
DT
XX
XX PCR primer for PST phosphatase interacting protein coding sequence.
DE
XX
XX PST phosphatase interacting protein; PSTPIP; tumour therapy;
KM
XX
XX protein tyrosine phosphatase; PCR primer; ss.
XX
XX Mammalia.
OS
XX
XX US6040437-A.
PN
XX
XX 21-MAR-2000.
PD
XX
XX 29-SEP-1997; 97US-0938830.
PF
XX
XX 07-FEB-1997; 97US-0104590.
PR
XX
XX (GETH) GENENTECH INC.
PA
XX
XX Dowbenko DJ, Lasky LA;
PI
XX
XX WPI; 2000-282393/24.
DR
XX
XX Novel genes encoding protein tyrosine phosphatase binding proteins
PT useful for isolating homologous genes, e.g. in tumour cells, which
PT provide more specific targets for tumour therapy
XX
XX Example 6; Column 43-44; 65pp; English.
XX
CC This sequence represents a PCR primer used to isolate the PST phosphatase
CC interacting protein (PSTPIP) coding sequence of the invention. The
CC protein is a protein tyrosine phosphatase that possesses a non-catalytic
CC domain comprising a proline, serine and threonine rich region and a

CC C-terminal segment of 20 amino acid (aa's) rich in proline, and defines
CC an SH3 binding domain. Nucleic acids encoding native PSTPIP molecules can
CC be used to isolate homologous genes specifically expressed in tumour
CC cells, which might provide more specific targets for tumour therapy. The
CC DNA is also useful for the preparation of PSTPIP polypeptides by
CC recombinant techniques and as hybridisation probes for searching cDNA and
CC genomic libraries for the coding sequence of other PSTPIP polypeptide
CC analogues in other species.
XX
XX SQ Sequence 47 BP; 9 A; 14 C; 15 G; 9 T; 0 other;
Alignment Scores:
Pred. No.: 0.135 Length: 47
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.41% Indels: 0
DB: 21 Gaps: 0
US-09-068-377C-1 (1-415) x AAA07303 (1-47)
QY 91 HisIIegInLeuAlaLeuAlaLeuArgGlu 100
DB 47 CACATCCAGCTGGCCCTGGCCTGCTAG 18
RESULT 13
AAV57984/C
ID AAV57984 standard; DNA; 48 BP.
XX
XX AAV57984;
AC
XX
XX 24-NOV-1998 (first entry)
DT
XX
XX Murine PSTPIP C-terminal deletion mutagenesis PCR primer C-coil.4.
DE
XX
XX Mouse; tyrosine phosphorylated cleavage furrow-associated protein;
KM PSTPIP; PEST family; protein tyrosine phosphatase; murine;
KM polymerisation; actin monomer; eukaryotic cell; identification;
KM antagonist; PCR primer; deletion mutagenesis; ss.
XX
XX Synthetic.
OS
XX
XX Mus sp.
OS
XX
XX WO9835037-A1.
PN
XX
XX 13-AUG-1998.
PD
XX
XX 30-JAN-1998; 98WO-US01774.
PF
XX
XX 29-SEP-1997; 97US-0938829.
PR
XX
XX 07-FEB-1997; 97US-0798419.
PA
XX
XX (GETH) GENENTECH INC.
PA
XX
XX Dowbenko DJ, Lasky LA;
PI
XX
XX WPI; 1998-447234/38.
DR
XX
XX New PEST-type protein tyrosine phosphatase interacting polypeptide -
PT nucleic acids and vectors; for inducing the polymerisation of actin
PT monomers in eukaryotic cells and identifying antagonists
XX
XX Example 6; Page 44; 11pp; English.
XX
CC The present sequence represents a PCR primer involved in deletion
CC mutagenesis of murine tyrosine phosphorylated cleavage furrow-associated
CC protein (PSTPIP), which is a PEST-type protein tyrosine phosphatase
CC (PTP)-interacting polypeptide. PSTPIP induces the polymerisation of
CC actin monomers in a eukaryotic cell, by introducing a vector containing
CC the nucleic acid sequence encoding PSTPIP into the cell. Assays for
CC identifying (antagonists of PSTPIP comprise contacting PSTPIP with the
CC agent and monitoring the ability of PSTPIP to induce actin
CC polymerisation.

XX Sequence 48 BP; 11 A; 18 C; 8 G; 11 T; 0 other;
SQ

Alignment Scores:
Pred. No.: 0.138 Length: 48
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.41% Indels: 0
DB: 19 Gaps: 0

US-09-068-377c-1 (1-415) x AAV57985 (1-48)

QY 255 ThrLeuGluglCysaspValgluglYasp 264
DB 47 ACCCTTGAGGCGCTGTGATGTGAAGGTGAC 18

RESULT 14
AAV57985/c
ID AAV57985 standard; DNA: 48 BP.
AC AAV57985;
XX
XX 24-NOV-1998 (first entry)
XX
XX Murine PSTPIP C-terminal deletion mutagenesis PCR primer C-coll.5.
DE
XX Mouse: tyrosine phosphorylated cleavage furrow-associated protein;
KM PSTPIP; PST family: protein tyrosine phosphatase; murine;
KM polymerisation; actin monomer; eukaryotic cell; Identification;
XX antagonist; PCR primer; deletion mutagenesis; ss.
XX
XX Synthetic.
OS Mus sp.
XX
XX W09835037-A1.
PN
XX 13-AUG-1998.
PD
XX 30-JAN-1998; 98MO-US01774.
PF
XX 29-SEP-1997; 97US-0938829.
PR 07-FEB-1997; 97US-0798419.
XX
XX (GETH) GENENTECH INC.
PA
XX Dowbenko DJ, Lasky LA;
PI
XX WPI; 1998-447234/38.
DR
XX
XX New PEST-type protein tyrosine phosphatase interacting polypeptide -
PT nucleic acids and vectors; for inducing the polymerisation of actin
PT monomers in eukaryotic cells and identifying antagonists
XX
XX Example 6; Page 44; 11pp; English.
PS
XX The present sequence represents a PCR primer involved in deletion
CC mutagenesis of murine tyrosine phosphorylated cleavage furrow-associated
CC protein (PSTPIP), which is a PST-type protein tyrosine phosphatase
CC (PTP)-interacting polypeptide. PSTPIP induces the polymerisation of
CC actin monomers in a eukaryotic cell, by introducing a vector containing
CC the nucleic acid sequence encoding PSTPIP into the cell. Assays for
CC identifying (ant)agonists of PSTPIP comprise contacting PSTPIP with the
CC agent and monitoring the ability of PSTPIP to induce actin
CC polymerisation.
XX
XX Sequence 48 BP; 12 A; 13 C; 14 G; 9 T; 0 other;
SQ

Alignment Scores:
Pred. No.: 0.138 Length: 48
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

Query Match: 2.41% Indels: 0
DB: 19 Gaps: 0

US-09-068-377c-1 (1-415) x AAV57985 (1-48)

QY 230 AlaLeuTryValHisCysasnglnLeuSer 239
DB 47 GCCCTTGAGGCGCTGTGATGTGAAGGTGAC 18

RESULT 15
AAA07301/c
ID AAA07301 standard; DNA: 48 BP.
XX
XX AAA07301;
AC
XX 27-JUN-2000 (first entry)
DT
XX PCR primer for PST phosphatase interacting protein coding sequence.
DE
XX PST phosphatase interacting protein; PSTPIP; tumour therapy;
KM protein tyrosine phosphatase; PCR primer; ss.
XX
XX Mammalia.
OS
XX US6040437-A.
PN
XX 21-MAR-2000.
PD
XX 29-SEP-1997; 97US-0938830.
PF
XX 07-FEB-1997; 97US-0104590.
PR
XX (GETH) GENENTECH INC.
PA
XX Dowbenko DJ, Lasky LA;
PI
XX WPI; 2000-282393/24.
DR
XX
XX Novel genes encoding protein tyrosine phosphatase binding proteins
PT useful for isolating homologous genes, e.g. in tumour cells, which
PT provide more specific targets for tumour therapy
XX
XX Example 6; Column 43-44; 65pp; English.
PS
XX This sequence represents a PCR primer used to isolate the PST phosphatase
CC interacting protein (PSTPIP) coding sequence of the invention. The
CC protein is a protein tyrosine phosphatase that possesses a non-catalytic
CC domain comprising a proline, serine and threonine rich region and a
CC C-terminal segment of 20 amino acid (aa's) rich in proline, and defines
CC an SH3 binding domain. Nucleic acids encoding native PSTPIP molecules can
CC be used to isolate homologous genes specifically expressed in tumour
CC cells, which might provide more specific targets for tumour therapy. The
CC DNA is also useful for the preparation of PSTPIP polypeptides by
CC recombinant techniques and as hybridisation probes for searching cDNA and
CC genomic libraries for the coding sequence of other PSTPIP polypeptide
CC analogues in other species.
XX
XX Sequence 48 BP; 11 A; 18 C; 8 G; 11 T; 0 other;
SQ

Alignment Scores:
Pred. No.: 0.138 Length: 48
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.41% Indels: 0
DB: 21 Gaps: 0

US-09-068-377c-1 (1-415) x AAA07301 (1-48)

QY 255 ThrLeuGluglCysaspValgluglYasp 264
DB 47 ACCCTTGAGGCGCTGTGATGTGAAGGTGAC 18

Mon Aug 4 09:40:06 2003

us-09-068-377c-1.olip2n.rng

Page 11

Search completed: August 3, 2003, 15:05:33
Job time : 329 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 3, 2003, 14:38:12; Search time 1852 Seconds

(without alignments)
3629.118 Million cell updates/sec

Title: US-09-068-377c-1

Perfect score: 415

Sequence: 1 MAAQLQFRDAFWCRDFTAH.....WTVERNGRGVFGSYLEKL 415

Scoring table:

OLIGO
Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 16154066 seqs, 809774376 residues

Word size: 1

Total number of hits satisfying chosen parameters: 32308016

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+pn.model -DEV=xlh
-O/cgcn2.1/uspto.spool/US09068377/runat_28072003_091403_15923/app.query.fasta.1.583
-DB=EST-QFMT-fstlap -SUFFIX=olip2n.rst -MINMATCH=0.1 -DOOPCL=0 -LOOPEXT=0
-UNITS-bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -DIST=45
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto
-NORM-ext -HEAPSIZE=500 -MINLEN=200000000
-USER-US09068377.qcgn.1.1.1906.etrnat.28072003.091403.15923 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEMEMORY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estin: *
4: em_estnu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_hic: *
9: gb_est1: *
10: gb_est2: *
11: gb_hic: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pin: *
21: em_gss_vtl: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_tod: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	226	54.5	888	13	B1556340
2	198	47.7	594	14	B0533004
3	117	28.2	629	12	BE853802
C 4	114	27.5	542	9	A1324866
C 5	114	27.5	568	13	BM221823
C 6	114	27.5	584	12	BF020734
7	112	27.0	502	9	AA038079
8	93	23.9	766	13	B1904968
9	93	22.4	495	10	AA916545
10	91	21.9	361	9	AA980931
11	88	21.2	423	12	BF019692
12	88	21.2	507	9	A1322422
13	81	19.5	414	14	BQ553003
C 14	78	18.8	440	13	BM222042
15	69	16.6	543	14	BM933148
16	61	14.7	560	10	AV616620
17	60	14.5	355	12	BF555456
18	60	14.5	447	12	BF549426
19	58	14.0	712	10	BB162096
C 20	56	13.5	401	13	B1290584
C 21	56	13.5	420	9	AA900024
C 22	56	13.5	455	13	B1290159
C 23	56	13.5	472	9	AA893267
C 24	56	13.5	518	14	B0190642
C 25	56	13.5	653	12	BF282676
C 26	56	13.5	733	14	BQ207763
C 27	54	13.0	547	10	AM540915
C 28	46	11.1	424	13	B1133449
29	45	10.8	420	9	AA866798
30	45	10.8	492	10	BE631504
31	35	8.4	512	13	BM087968
32	35	8.4	521	12	BF073935
33	32	7.7	370	9	A1877165
34	31	7.5	381	13	B1029004
35	31	7.5	483	10	AM659798
36	31	7.5	572	10	AM657869
37	31	7.5	797	12	BG820623
38	31	7.5	828	12	BG819706
39	31	7.5	1209	13	BM547486
40	30	7.2	326	10	BM459563
41	30	7.2	697	9	AL559399
42	30	7.2	698	12	BG676651
43	30	7.2	707	13	B1906497
44	30	7.2	708	13	B1906059
45	30	7.2	737	13	B1910444

ALIGNMENTS

RESULT 1
LOCUS B1556340
DEFINITION B1556340 888 bp mRNA
LOCUS 603237461F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5290479 5', mRNA sequence.
ACCESSION B1556340
VERSION B1556340.1 GI:15443654
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
AUTHORS I (bases 1 to 888)
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL
COMMENT

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MSC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM11734 row: 3 column: 16
High quality sequence stop: 783.

FEATURES
source

Location/Qualifiers
1..888
/organism="Mus musculus"
/strain="129,C57BL/6J,FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5290479"
/clone_lib="NCLCGAP_Mam3"
/tissue_type="tumor, gross tissue"
/dev_stage="10 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: PCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference for transgenic model: Xu et al., Nature Genetics 22, 37-43 (1999)."
BASE COUNT 241 a 227 c 263 g 157 t
ORIGIN

Alignment Scores:

Pred. No.: 3,87e-213 Length: 888
Score: 226.00 Matches: 226
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 54.46% Indels: 0
DB: 13 Gaps: 0

US-09-068-377c-1 (1-415) x B1556340 (1-888)

QY 85 GUAAGNVAIGYSEALAHISLEGINLEUALALEUALAUAARGUGLUGLUAARGSER 104
|||||
DB 2 GAGAAATGTGGCGGTGCACACATCCAGCTGGCCCTGGCTGTGAGAGCTGGGAGC 61
QY 105 LEUGLUGLUPHEARGLUARGGLNLYSGULNARGLYSTYRGUAAIAIEMETASP 124
|||||
DB 62 CTGGAGAGATTCCGAGAGACAGAAAGACCGGAGAAAGTATGAGGCCATCATGAC 121
QY 125 ARGVALGILNLYSERLYSEUSERTLYRLYSTHMETGLUSERLYSLYSALATYR 144
|||||
DB 122 CGTGTCCAGAGAGCAAGTTGCTGTACAAAGACCATGGATGCCAAGAGGCATAT 181
QY 145 ASPGLILYSCYARGPSALAAASPAPALAGLUGLUALAHPHEGLUARVALSERIALA 164
|||||
DB 182 GACCAAGATGCGAGGATGATGATGCTGACAGAGCCCTTCGAGCTGTGAGGCCAAT 241
QY 165 GLYHISGLINLYSGINVALGILNLYSERGLNASNLYSALALYSGINCYSLYSGIUSERALA 184
|||||
DB 242 GGCACACAGAACCAAGTAGAAAAGAGCCAGAACCAAGCAGTCAAGAGATCGCC 301
QY 185 THHGLUALAGLUARGVALTYRARGLNASHILLEGULINLEUGLNUARGALATGTHGLU 204
|||||
DB 302 ACAGAGAGCAGAAAGAGTACAGGCAAAATATCGAACATCTCGAGAGAGCAGACCGAG 361
QY 205 TRPGLUGLNUHISARGTHRTYRTHRCYSGUALAHPHEGLINLEUGLNUHPHASPARGLEU 224
|||||
DB 362 TGGAGAGAGAGACACCGGACTACTGTGAGCGCTTCAGTTGCGAGATTGACCGGCTC 421
QY 225 THRTLEUARGSNALALEUTTPVALHISCYASNGLINLEUSERMETGINCYSVALLYS 244
|||||
DB 422 ACCATCTCCGCAATGCGCTGTGGTGCTACTGTAAACAGCTCTCCATGACAGTGTGTAAG 481

QY 245 ASPASGLULEUTYRCUGLUGLUALAARGLEUTHLEUGLUGLYCYASAPVALUGLYASP 264
|||||
DB 482 GATGATGACCTTATGAGCAAGTGCAGCTGACCCCTTGAGCGCTGTATGTGAGAGCTGAC 541
QY 265 ILEASNGLYPHEILLEGINSERLYSERTTHRGYARGGLUPROVALPROVALPROTYR 284
|||||
DB 542 ATCAATGGCTTCATCCAGTCCAGCAAGAGACTGGCAGAGAGCCCCCAGCTCGGCTTAT 601
QY 285 GLNASRTTYTARSPARGLUVALTHRPROLEILLEGLYSERPROSERILEGINSER 304
|||||
DB 602 CAGAACTACTATGACAGGAGGAGTGCACCTGATTGGACCCCTAGCATCCAGCCCTCC 661
QY 305 CYSGLYVALILEYSARG 310
|||||
DB 662 TCGGCTGTGATAAAGAGC 679

RESULT 2
B0553004 594 bp mRNA linear EST 20-JUN-2002
LOCUS H4019E01-5, mRNA sequence.
DEFINITION H4019E01 5', mRNA sequence.
ACCESSION B0553004 GI:21453892
VERSION B0553004
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS

1 (bases 1 to 594)
VanBuren,V., Piao,Y., Dudekula,D.B., Qiao,Y., Carter,M.G., Martin
P.R., Stagg,C.A., Bassey,U., Alba,K., Hamatani,T., Kargul,G.J.,
Luo,A.G. and Ko,M.S.H.
Assembly, verification, and initial annotation of NIA 7.4K mouse

TITLE
JOURNAL

CDNA clone set
Unpublished (2002)
Other_ESTs: H4019E01-3

COMMENT

Contact: Yong Qian
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 3000, Baltimore, MD 21224-6820, USA
Email: cnaelgusn.grc.nia.nih.gov
This clone set has been freely distributed to the community. Please
visit http://lgusn.grc.nia.nih.gov/cDNA/NIA_7_4K.html for details.
Plate: H4019 row: E column: 01
Seq primer: -21M13 Reverse
High quality sequence stop: 594
POLYA-No.

FEATURES

source

Location/Qualifiers
1..594
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:H4019E01-5"
/db_xref="taxon:10090"
/clone="H4019E01"
/clone_lib="NIA Mouse 7.4K cDNA Clone Set"
/sex="mixed"
/dev_stage="mixed"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
clone is among a rearrayed set of 7,407 clones from more
than 20 cDNA libraries."
BASE COUNT 177 a 130 c 207 g 80 t
ORIGIN

Alignment Scores:

Pred. No.: 1.32e-185 Length: 594
Score: 198.00 Matches: 198
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 47.71% Indels: 0
DB: 14 Gaps: 0

US-09-068-377c-1 (1-415) x B0553004 (1-594)

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 542)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dierich, N., Dubuque, T.,
Gelsel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through INL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:283739
This clone was previously sequenced on the 5' end only, this new
data is from the 3' end
Possible reversed clone: similarity on wrong strand
High quality sequence stop: 473.
Location/Qualifiers
1. 542
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:472995"
/clone_lib="Soares mouse p3NM19.5"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCACATCTGAGTGGAGCGCGCGCTCTTCTTTTCTTTTCTTTT 3'],
TGTTCACATCTGAGTGGAGCGCGCGCTCTTCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University)."
BASE COUNT 112 a 152 c 155 g 123 t
ORIGIN
Alignment Scores:
Pred. No.: 1.81e-102 Length: 542
Score: 114.00 Matches: 114
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 27.47% Indels: 0
DB: 9 Gaps: 0
US-09-068-377c-1 (1-415) x A1324866 (1-542)
QY 302 GlnProSerCysGlyValIleuLysArgPheSerGlyLeuLeuHisGlySerProLysThr 321
Db 524 CAGCCCTCTGCTGTGATTAAGAGCTTCTGGGCTGTACATGGAAGTCCCAAGACC 465
QY 322 ThrProSerAlaProAlaIleSerThrGluThrLeuThrProThrProGluArgAsnGlu 341
Db 464 ACACCTTCTGCTCTCTGCTTCCACAGACACTGTGACTCCACCCCTGAGCGGANTGAG 405
QY 342 LeuValTyrAlaSerIleGluValGlnAlaThrGlnGlyAsnLeuAsnSerSerAlaGln 361
Db 404 TTGGCTACGACCATTCGAAAGTCAGGCGACCCAGGAGAACTTAACATCATCAGCCAG 345
QY 362 AspTyrArgAlaLeuTyrAspTyrThrAlaGlnAsnSerSpGluLeuAspIleSerAla 381
Db 344 GACTACGGGCGACTTACGACTACCGCGACAGAAATTTGTATGAGTGCAGCATTTCCGCG 285

QY 382 GlyAspIleLeuAlaValIleuGluGlyIleuAspGlyTyrTrpThrValGluArgAsn 401
Db 284 GGAGACATCTCGCGGTCTCTCGAAGGGAGATGCTGTGACACTGTGAGCGGAGAC 225
QY 402 GlyGlnArgGlyPheValProGlySerTyrLeuGluLysLeu 415
Db 224 GGACACAGTGGCTTTGTCTCGTGGTCTGAGTCTGAGAAACCTC 183
RESULT 5
BM221823/c
LOCUS
DEFINITION K0110F02-3 NIA Mouse Hematopoietic Stem Cell (Lin-/c-Kit+/Sca-1-)
CDNA Library (Long) Mus musculus cDNA clone K0110F02 3', mRNA
sequence.
ACCESSION BM221823
VERSION BM221823.1 GI:17782294
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 568)
AUTHORS Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Luo, A., Carter, M.G.,
Taub, D., Longo, D.L., Keller, J. and Ko, M.S.H.
TITLE Systematic Analyses of NIA Mouse Hematopoietic Stem Cell
(Lin-/c-Kit+/Sca-1-) cDNA Library (Long)
JOURNAL Unpublished (2001)
COMMENT Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdaa@nigun.grc.nia.nih.gov
Plate: K0110 row: F column: 02
Seq primer: -21M13 Forward
High quality sequence stop: 568
POLYA=Yes.
Location/Qualifiers
1. 568
/organism="Mus musculus"
/strain="C57BL/6NCR"
/db_xref="niaEST:K0110F02-3"
/db_xref="taxon:10090"
/clone="K0110F02"
/clone_lib="NIA Mouse Hematopoietic Stem Cell
(Lin-/c-Kit+/Sca-1-) cDNA Library (Long)"
/tissue_type="Hematopoietic Stem Cell (Lin-/c-Kit+/Sca-1-
)"
/dev_stage="Age approx. 10 weeks old"
/lab_host="DH10B"
/note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
obtained from Drs. Dennis Taub, Dan Longo (National Cancer
Institute, USA). Double-stranded cDNAs were synthesized
with an oligo(dT) primer [Invitrogen]:
5'-pGACTACTTACATCGCGAGCGCGCCCTTTTCTTTTCTTTT-3' from
2.4 ug of total RNA, treated with T4 DNA polymerase, and
purified by ethanol-precipitation. The cDNAs were ligated
to lone-linker L1-SalI, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer Sal4-S. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pSPORT1 plasmid vector.
The DH10B E. coli host was transformed with the ligation
mixture by the standard chemical method. The average
insert size is about 2.2 kb. The library was constructed
by Tuian Piao (NIA)."

BASE COUNT 116 a 162 c 164 g 126 t
ORIGIN

Alignment Scores:
Pred. No.: 1,96e-102 length: 568
Score: 114.00 Matches: 114
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 27.47% Indels: 0
DB: 13 Gaps: 0

US-09-068-377c-1 (1-415) x BM221823 (1-568)

QY 302 GlnProSerCysGlyValIleLysArgPheSerGlyLeuLeuHisGlySerProLysThr 321
|||
Db 512 CAGCCCTCTGCTGCTGATTAAGAGGTTCTCTGGGCTGTACATGAAAGTCCCAAGACC 453
|||

QY 322 ThrProSerAlaProAlaAlaSerThrGluThrLeuThrProThrProGluArgAsnGlu 341
|||
Db 452 ACACCTTCTGCTCTGCTGCTCCACAGAGACTGTGATCCACCCTGACCGGAGATGAG 393
|||

QY 342 LeuValTyraLaserIleGluValAlaIleThrGlnGlnLysAsnLeuAsnSerSerAlaGln 361
|||
Db 392 TTGGTCTACCGCATTCATCGAAGTGCAGCCGACGAAACCTTAATCTATCAGCCGAG 333
|||

QY 362 AspTyraArgAlaLeuTyraPtyrThrAlaGlnAsnSerAspGluLeuAspIleSerAla 381
|||
Db 332 GACTACCGGCGACACTCTACGACTACACCGACAGAAATTCGTGATGAGCTGACATTTCCGCG 273
|||

QY 382 GlyAspIleLeuAlaValIleLeuGluGlyLysAspGlyTyrPtyrThrValGluArgAsn 401
|||
Db 272 GAGACATCTCGCGGTCATCTGTAAGGAGAGATGCTGCTGATGCTGAGACGCGAAC 213
|||

QY 402 GlyGlnArgGlyPheValProGlySerTyrLeuGluLysLeu 415
|||
Db 212 GGACAACTGGCTTTGTCCCTGGCTGCTACTTGAGAGAAAGCTC 171
|||

RESULT 6
BF020734/c 584 bp mRNA linear EST 10-OCT-2000
LOCUS
DEFINITION
IMAGE:3472365.3, similar to TR:P97814 P97814 PEST PHOSPHATASE
INTERACTING PROTEIN, mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 584)
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Other-ESTs: ux23f11.y1
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:1391725
High quality sequence step: 380.
Location/Qualifiers
1. 584
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone IMAGE:3472365"
/clone_lib="Soares_thymus_2NBMT"
/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"

FEATURES
Source

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCACCATCTGGAAGTGGAGCGGCCGGCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. RNA
provided by Dr. Bertrand Jordan. Library went through two
rounds of normalization, and was constructed by Bento
Soares and M. Patricia Bonaldo."

BASE COUNT 119 a 163 c 165 g 137 t
ORIGIN

Alignment Scores:
Pred. No.: 1,96e-102 length: 584
Score: 114.00 Matches: 114
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 27.47% Indels: 0
DB: 12 Gaps: 0

US-09-068-377c-1 (1-415) x BF020734 (1-584)

QY 302 GlnProSerCysGlyValIleLysArgPheSerGlyLeuLeuHisGlySerProLysThr 321
|||
Db 518 CAGCCCTCTGCTGCTGATTAAGAGGTTCTCTGGGCTGTACATGAAAGTCCCAAGACC 459
|||

QY 322 ThrProSerAlaProAlaAlaSerThrGluThrLeuThrProThrProGluArgAsnGlu 341
|||
Db 458 ACACCTTCTGCTCTGCTGCTCCACAGAGACTGTGATCCACCCTGAGGAGATGAG 399
|||

QY 342 LeuValTyraLaserIleGluValAlaIleThrGlnGlnLysAsnLeuAsnSerSerAlaGln 361
|||
Db 398 TTGGTCTACCGCATTCATCGAAGTGCAGCCGACGAAACCTTAATCTATCAGCCGAG 339
|||

QY 362 AspTyraArgAlaLeuTyraPtyrThrAlaGlnAsnSerAspGluLeuAspIleSerAla 381
|||
Db 338 GACTACCGGCGACACTCTACGACTACACCGACAGAAATTCGTGATGAGCTGACATTTCCGCG 279
|||

QY 382 GlyAspIleLeuAlaValIleLeuGluGlyLysAspGlyTyrPtyrThrValGluArgAsn 401
|||
Db 278 GAGACATCTCGCGGTCATCTGTAAGGAGAGATGCTGCTGATGCTGAGACGCGAAC 219
|||

QY 402 GlyGlnArgGlyPheValProGlySerTyrLeuGluLysLeu 415
|||
Db 218 GGACAACTGGCTTTGTCCCTGGCTGCTACTTGAGAGAAAGCTC 177
|||

RESULT 7
AA038079 502 bp mRNA linear EST 28-AUG-1996
LOCUS
DEFINITION
IMAGE:472995.5, similar to SW:MYSC_ACACA P10569 MYOSIN IC HEAVY
CHAIN [1], mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus
house mouse.
EST.
AA038079.1 GI:1513518

REFERENCE
AUTHORS
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 502)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R., and
Waterston, R.
The WashU-HM Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HM Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800

TITLE
JOURNAL
COMMENT

Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:283739
 Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 494.
 Location/Qualifiers

FEATURES

source

1..502
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:472995"
 /clone_lib="Soares mouse p3NNF19.5"
 /dev_stage="19.5 dpc total fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Vector: p773D (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5'.
 TGTATCCATCTGAGTGGAGCGCGCATTTTCTTTTCTTTT 3']
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified p773 vector
 (Pharmacia). Library went through one round of
 normalization to a Cot = 5. Library constructed by Bento
 Soares and M. Fatima Bonaldo. RNA was kindly provided by
 Dr. Minoru Ko (Wayne State University)."
 BASE COUNT 112 a 140 c 147 g 103 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1,59e-100 Length: 502
 Score: 112.00 Matches: 166
 Percent Similarity: 98.81% Conservative: 0
 Best Local Similarity: 98.81% Mismatches: 1
 Query Match: 26.99% Indels: 2
 Gaps: 0

US-09-068-377C-1 (1-415) x AA038079 (1-502)

OY 247 GluLeuTYrGluGluValAlaArgLeuThrLeuGluGlyCysAspValGluGluAspIleasn 266
 |||||||
 DB 1 GAGCTCTATGAGGAGTGGCGCTGACCCCTTGAGGCGCTGATGTGAGAGGTATCATCACT 60
 OY 267 GlyPheIleGlnSerLysSerThrGlyArgGluProProAlaProValProTyrGlnasn 286
 |||||||
 DB 61 GCGTTCATCCAGTCCAGAGCACTGGCAGAGAGCCCCAGCTCCGCTGCTTATCAGAAC 120
 OY 287 TyrTyrAspArgGluValThrProLeuIleGlySerProSerIle-GlnProSerCysG 306
 |||||||
 DB 121 TACTATGACAGGAGGTGACCCCACTGATGGCAGCCCTAGCGGT-CCAGCCCTCTGTGG 179
 OY 306 ValAlleLysArpPheSerGlyLeuLeuHisGlySerProLysThrThrProSerAlaPr 326
 |||||||
 DB 180 TGTGATTAAGAGGTTCTCTGGGCTGCTACATGAGAGTCCCAAGACACACCTTCGCTCC 239
 OY 326 CalAlaSerThrGluThrLeuThrProThrProGluArgasnGluLeuValTyrAlaSe 346
 |||||||
 DB 240 TGCCTCTCCACAGAGACTGTGACTCCACCCCTAGCGGAGATGATGGTCTACGCAATC 299
 OY 346 rIleGluValGlnAlaThrGlnGlyAsnLeuasnSerSerAlaGlnAspTyrArgAlaLe 366
 |||||||
 DB 300 CATGAACTGCAGGCGGACCCAGGAAACCTTAACATCAGCCAGGACTACCGGCACT 359
 OY 366 uTyrAspTyrThrAlaGlnAsnSerAspGluLeuAspIleSerAlaGlyAspIleLeuAl 386
 |||||||
 DB 360 CTAGAGACTACACCGACAGATTCGATGAGCTGACATTCGCCGGGAGACATCTCGC 419
 OY 386 aValIleLeuGluGlyGluAspGlyTyrTyrThrValGluArgasnGlyGlnArgGlyPh 406
 |||||||
 DB 420 GGTGATCTCTGGAAGGGAGATGGCTGCTGATGCTGAGCGGAGCAACGAGCAACGTCGCT 479
 OY 406 eValProGlySerTyrLeuGlu 413
 |||||||

DB 480 TGTCTCTGGGTCTACTGTGAG 501

RESULT 8
 LOCUS BI904988 766 bp mRNA linear EST 16-OCT-2001
 DEFINITION 603169121F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5257060 5',
 mRNA sequence.
 ACCESSION BI904988
 VERSION BI904988.1 GI:16167418
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.

REFERENCE: NIH-MGC <http://mgc.nci.nih.gov/>.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgaps-remail.nih.gov
 Tissue procurement: Gilbert Smith, Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
 Plate: L1AM11649 row: j column: 05
 High quality sequence start: 6
 High quality sequence stop: 764.
 Location/Qualifiers

FEATURES

source

1..766
 /organism="Mus musculus"
 /strain="C57BL/6 J"
 /db_xref="taxon:10090"
 /clone="IMAGE:5257060"
 /clone_lib="NCI_CGAP_Lu33"
 /tissue_type="pooled lung tumors"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: Lung; Vector: p773D-Pac (Pharmacia) with a
 modified polylinker. Site_1: NotI; Site_2: EcoRI; 1st
 strand cDNA was prepared from mRNA obtained from pooled
 lung tumors with a Not I - oligo(dT) primer [5'.
 TGTATCCATCTGAGTGGAGCGCGCATTTTCTTTTCTTTT 3']
 double-stranded cDNA was ligated to Eco RI adapters
 (Pharmacia), digested with Not I and cloned into the Not
 I and Eco RI sites of the modified p773 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 165 a 217 c 220 g 164 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1.88e-87 Length: 766
 Score: 99.00 Matches: 173
 Percent Similarity: 98.86% Conservative: 0
 Best Local Similarity: 98.86% Mismatches: 1
 Query Match: 23.86% Indels: 2
 Gaps: 0

US-09-068-377C-1 (1-415) x BI904988 (1-766)

OY 242 CysValLysAspAspGluLeuTyrGluValAlaArgLeuThrLeuGluGlyCysAspVal 261
 |||||||
 DB 90 TGTGTCAGAGATGATGAGCTCTATGAGGAGTGGCGCTGACCCCTTGAGGCGCTGATGTG 149
 OY 262 GluGlyAspIleasnGlyPheIleGlnSerLysSerThrGlyArgGluProProAlaPro 281
 |||||||
 DB 150 GAAGTGCATCAATGAGCTTATCAGTCACCAAGACACTGGAGAGACCCCACTCGC 209
 OY 282 ValProTyrGlnAsnTyrTyrAspArgGluValThrProLeuIleGlySerProSerIle 301
 |||||||

Db	210	GTGGCTTATACAGACTACTAATGACAGAGGAGGTGAACCCACATCGATTGGAGGCCATCAGATC	265
Oy	302	GlnProSerCySgIlyValIleLysArgPheSerGlyLeuHisGlySer-ProLysThr	321
Db	270	CAGCCCTTCCTCGCGTGATTAAGAGGTTCTCTGGGCTGCATACTGGAAGTCCCAAGACC	329
Oy	322	ThrProSerAlaProAlaAlaSerThcGluThrLeuthrProThr-ProtGuarGAsnGl	341
Db	330	ACACCTTCCTGCTCGCTGCCTTCCACAAGAAGACTGTGATCTCCACGCCCTGAGCGGAAGA	389
Oy	341	IleuValTyrrAlaSerIleGluValGlnAlaThrcGlnLysAsnLeuAsnSerSerAlaGl	361
Db	390	GTTGGTCTACGCATCATCCAGAAAGTCAGAGCCACCCA-GGAACCTTAACTATATAGCCCCA	448
Oy	361	nASPtyrArGalaleuTyraSPtyrThrAlaGlnAsnSeraspGluLeuaspIleSerAl	381
Db	449	GGACTACGGGACACTCTFACGACTACACTGCACAGAAATTCGATGAGCTGGACATTTCCGC	508
Oy	381	aGIYasPIleLeuAlaValIleLeuGluGlyGluAspGlyTrpThrValGluArGas	401
Db	509	GGGAGACATCTCTGGGGGTCACTCCGGAAGGGAGAGATGGCTGAGTGGAGCGGAG	568
Oy	401	nglYglnaRgIlypheValProGlyserTYrLeuGlnLysLeu	415
Db	569	CGGACAACTGTCGCTTGTCTCCGTGGTCTACTTGGAGAACTC	611
RESULT 9			
AM916545			
LOCUS	AM916545	495 bp	mRNA linear EST 25-MAY-2000
DEFINITION	ESR347849 Rat gene index, normalized rat, norvegicus, Bento Soares		
ACCESSION	Rattus norvegicus cDNA clone RGD1956 5' end, mRNA sequence.		
VERSION	AM916545		
KEYWORDS	AM916545..1 GI:8082273		
SOURCE	EST.		
ORGANISM	Norway rat.		
	Rattus norvegicus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;		
	Rattus.		
REFERENCE	1 (bases 1 to 495)		
AUTHORS	Lee,N.H., Glodex,A., Chandra,I., Mason,T.M., Quackenbush,J.,		
TITLE	Kerlavage,A.R. and Adams,M.D.		
JOURNAL	Rat Genome Project: Generation of a Rat EST (REBT) Catalog & Rat		
COMMENT	Gene Index Unpublished (1998) Contact: Lee, NH The Institute for Genomic Research 9712, Medical Center Drive, Rockville, MD 20850, USA Tel.: (301)-838-3529 Fax: (301)-838-0208 Email: nhlee@tigr.org This clone is available through the ATCC, contact the ATCC tel#703-365-2700 for further information Seq primer: M13 Reverse. Location/Qualifiers 1..495 /organism="Rattus norvegicus" /db_xref="taxon:10116" /clone="RGIDR56" /clone_lib="Rat gene index, normalized rat, norvegicus, Bento Soares" /tissue_type="mix - brain, ovary, placenta, kidney, lung, liver, embryo, heart, muscle, spleen" /lab_host="SOLR" /note="vector: pluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Estimated insert size approx.1 kb"		
BASE COUNT	126 a 116 c 170 g 83 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	1.02e-81	Length:	495
Score:	93.00	Matches:	93
Percent Similarity:	100.00%	Conservative:	0

Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	22.41%	Indels:	0
DB:	10	Gaps:	0

US-09-068-377C-1 (1-415) x AW916545 (1-495)

OY	13	CysatrgaspphethrAlaHisThhGlyTyrGluValLeuLeuGlnArgLeuAspGly	32
Db	82	TGCAGGGACCTTCACGGCCACACAGGGATATAGTACTCTCGACAGAGCTGTGGACGGC	141
OY	33	ArglysmetCysLysAspValGluGluLeuLeuArgGlnArgAlaGlnAlaGluArg	52
Db	142	AGGAAGATGTCCAAAGCATGTGGAGAGCTGCTCAGACAGAGGCCCGAGCGAGAGAGC	201
OY	53	TyrGlyLysGluLeuValGlnIleAlaArgLysAlaGlyGlyGlnThrGluMetAsnSer	72
Db	202	TACGGGAAGAGCGTGGTGCACATTTGGCGGCAACACTGGTGGCCAGACAGATGAATTC	261
OY	73	LeuArgThrSerPheAspSerLeuLysGlnGlnThrGluAsnValGlySerAlaHisIle	92
Db	262	CTGAGGACCTCTTGTGACTCTCTGACAGCAGCAACAGAAATGTGGCGACAGCATC	321
OY	93	GlnLeuAlaLeuAlaLeuArgGluGluLeuArgSerLeu	105
Db	322	CAGCTGGCCCTGGCCCTGCTGTGAGGAGCTGGGAGCCTG	360

RESULT 10
LOCUS AA980931
DEFINITION u047e05.t1 Soares_mammary_gland_NbMXMg Mus musculus cDNA clone IMAGE:1349888 5' similar to TR:P97814 P97814 PEST PHOSPHATASE INTERACTING PROTEIN.; mRNA sequence.
ACCESSION AA980931
VERSION AA980931.1 GI:3159467
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 361)
Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisels, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
The Mashu-HHMI Mouse EST Project
Unpublished (1996)
Contact: Marras M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:698680
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 276.
Location/Qualifiers
1..361
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1349888"
/clone_id="Soares_mammary_gland_NbMXMg"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - Oligo(dT)

primer [5'
TGTACCAATCTGAAGTGGAGCGCCGCAATGTTTTTTTTTTTTTTTTTT
T 3']: double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified p773 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M. Fatima
Bonaldo."

BASE COUNT 87 a 90 c 108 g 76 t
ORIGIN

Alignment Scores:
Pred. No.: 6,88e-80 Length: 361
Score: 91.00 Matches: 91
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 21.93% Indels: 0
DB: 9 Gaps: 0

US-09-068-377C-1 (1-415) x AA980931 (1-361)

QY 237 GlnleuserMetGlnCysValIysaspGluLeuTyrGluValArgLeuThrIleu 256
|||||
DB 1 CAGCTCTCAGTCAAGTGTCTCAAGATGATGACTTATGAGGAGAGTGGCTGACCTT 60
|||||
QY 257 GlnGlyCysAspValGluGlyAspIleasnGlyPheIleGlnSerTyrGlyArg 276
|||||
DB 61 GAGGGCTGTGATGGAGAGTGAACATCAATGCTTCATCCAGACAGACAGTGGCAGA 120
|||||
QY 277 GluProPAlaProValProTyrGlnAsnTyrTyrAspArgGluValThrProLeuIle 296
|||||
DB 121 GAGCCCCAGCTCCGGTCCCTTATCAGAACTATGACAGGAGGCTGACCCCACTGATT 180
|||||
QY 297 GlySerProSerIleGlnProSerCysGlyValIleValAspPheSerGlyLeuLeuHis 316
|||||
DB 181 GGCACCTCAGTACATCCAGCCCTCTCCGCTGTGATTAAGAGTTCTCTGGCTGCTACAT 240
|||||
QY 317 GlySerProLysThrThrProSerAlaProAla 327
|||||
DB 241 GGAAGTCCCAAGACCACACCTTCTGCTCTGCT 273
|||||

RESULT 11
LOCUS BF019692 423 bp mRNA linear EST 10-OCT-2000
DEFINITION uxi4g08.y1 Soares_thymus_2NBMT Mus musculus cDNA clone
IMAGE:3471518 5' similar to TR:P97814 P97814 PEST PHOSPHATASE
INTERACTING PROTEIN. ; mRNA sequence.
ACCESSION BF019692
VERSION BF019692.1 GI:10751024
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 423)
REFERENCE NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
TITLE Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-rt@mail.nih.gov
This clone is available royalty-free through LNC; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1390878
Seq primer: -40RP from GIBCO
High quality sequence stop: 369.
Location/Qualifiers
1..423
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3471518"
/clone_lib="Soares_thymus_2NBMT"

/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAAGTGGAGCGCCGCAATGTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p773 vector. RNA
provided by Dr. Bertrand Jordan. Library went through two
rounds of normalization, and was constructed by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT 97 a 119 c 120 g 87 t
ORIGIN

Alignment Scores:
Pred. No.: 7.65e-77 Length: 423
Score: 88.00 Matches: 88
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 21.20% Indels: 0
DB: 12 Gaps: 0

US-09-068-377C-1 (1-415) x BF019692 (1-423)

QY 328 AlaSerThrGluThrLeuThrProThrProGluArgAsnGluLeuValTyrAlaSerIle 347
|||||
DB 73 GCTTCCACAGAGACTGTGATCCACCCCGAGCGGAGTATGTTGGTTCGATCCATCC 132
|||||
QY 348 GluValAlaIlnAlaThrGlnGlyAsnLeuAsnSerSerAlaGlnAspTyrArgAlaLeuTyr 367
|||||
DB 133 GAATGAGAGCGGAGCCAGGAACCTTAATCTACACCCAGAGACTACCGGACACTCTAC 192
|||||
QY 368 AspTyrThrAlaGlnAsnSerAspGluLeuAspIleSerAlaGlyAspIleAlaVal 387
|||||
DB 193 GACTACACCCACAGATTTCTGATGAGCTGACATTTCCCGGAGAGACATCTGGCGGTC 252
|||||
QY 388 IleLeuGluGlyGluAspGlyTyrThrThrValGluArgAsnGlyGlnArgGlyPheVal 407
|||||
DB 253 ATCTGGAGAGGGAGAGTGGCTGTGATGAGCTGTGAGAGCGAGACGACACTGCTTGTTC 312
|||||
QY 408 ProGlySerTyrLeuGluLysLeu 415
|||||
DB 313 CCTGGCTGTCTACTTGGAGAGCTC 336
|||||

RESULT 12
LOCUS A1322422 507 bp mRNA linear EST 23-DEC-1998
DEFINITION m181d02.y1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
IMAGE:472995 5' similar to TR:P97814 P97814 PEST PHOSPHATASE
INTERACTING PROTEIN. ; mRNA sequence.
ACCESSION A1322422
VERSION A1322422.1 GI:4056851
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 507)
REFERENCE Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Maria M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800


```

QY      395  TTTPrThrValaGlnArGAnGlyGlnArgGlyPheValProGlySerTyrLeuGlnLys 414
      |||||||
Db      233  TGGTGGACTGTGGAGCGGAGACGACACACGTGGCTTTGCTCCTGGCTGACTTGAGACAG 174

QY      415  Leu 415
      |||
Db      173  CTC 171

RESULT 14
BM222042/c 440 bp  mRNA  linear  EST 31-JAN-2002
LOCUS      K0114B02-3 NIA Mouse Hematopoietic Stem Cell (Lin-/c-Kit+/Sca-1-)
DEFINITION cDNA library (long) Mus musculus cDNA clone K0114B02 3', mRNA
sequence.
ACCESSION  BM222042
VERSION     BM222042.1  GI:17782552
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
REFERENCE  1 (bases 1 to 440)
AUTHORS   Piao,Y., Kargul,G.J., Dudekula,D.B., Qian,Y., Luo,A., Carter,M.G.,
          Taub,D., Longo,D.L., Keller,J. and Ko,M.S.H.
          Systematic Analyses of NIA Mouse Hematopoietic Stem Cell
          (Lin-/c-Kit+/Sca-1-) cDNA Library (long)
          Unpublished (2001)
JOURNAL   Contact: Dawood B. Dudekula
          Laboratory of Genetics
          National Institute on Aging/National Institutes of Health
          333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
          Email: cdnaelg@sun-grc.nia.nih.gov
          Plate: K0114 row: B column: 02
          Seq primer: -21M13 Forward
          High quality sequence stop: 440
          POLY-A-Yes.
FEATURES
source     Location/Qualifiers
1..440
/organism="Mus musculus"
/strain="C57BL/6NCr"
/db_xref="nleST:K0114B02-3"
/db_xref="taxon:10090"
/clone="K0114B02"
/clone_1lb="NIA Mouse Hematopoietic Stem Cell
(Lin-/c-Kit+/Sca-1-) cDNA Library (long)"
/tissue_type="Hematopoietic Stem Cell (Lin-/c-Kit+/Sca-1-
)"
/dev_stage="Age approx. 10 weeks old"
/lab_host="DH10B"
/note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://igsun-grc.nia.nih.gov/cDNA). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001)). [PMID: 11544199]". Total Genome Res.
obtained from Drs. Dennis Taub, Dan Longo (National
Institute on Aging, USA), Jonathan Keller (National Cancer
Institute, USA). Double-stranded cDNAs were synthesized
with an Oligo(dT) primer (Invitrogen:
5'-pGACTGATCTGATCGAGCGCGCCCTCTTTTCTTTT-3') from
2.4 ug of total RNA, treated with T4 DNA polymerase, and
purified by ethanol-precipitation. The cDNAs were ligated
to lone-linker L1-SalI, purified by phenol/chloroform, and
separated from free linkers by Centurion 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer Sal14-S. The
products were purified by phenol/chloroform and Centurion
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pSPORT1 plasmid vector.
The DH10B E. coli host was transformed with the ligation
mixture by the standard chemical method. The average
insert size is about 2.2 kb. The library was constructed

```

```

BASE COUNT      86 a      130 c      127 g      97 t
ORIGIN
Alignment Scores:
Pred. No.:      6.38e-67      Length:      440
Score:          78.00      Matches:      78
Percent Similarity: 100.00%      Conservative: 0
Best local Similarity: 100.00%      Mismatches: 0
Query Match:     18.80%      Indels:      0
DB:              13      Gaps:      0

US-09-068-377c-1 (1-415) x BM222042 (1-440)

QY      326  ProAlaIAserThrGluThrLeuThrProThrProGluArGAnGlyLeuValTyrAla 345
      |||||||
Db      440  CCGCGTCTTCCACAGAGACTGTGACTCCACCCCTGAGCGGATGATGTTGCTACGCA 381

QY      346  SerIleGluValGlnAlaThrGlnGlyAsnLeuAsnSerSerAlaGlnAspTyrArgAla 365
      |||||||
Db      380  TCCATCGAAGTGCAGCGGACCCAGGAAACCTTAATCATCATCAGCCAGGACCTACCGGCA 321

QY      366  LeuTyrAspTyrThrAlaGlnAsnSerAspGluLeuAspIleSerAlaGlyAspIleLeu 385
      |||||||
Db      320  CTCTACGACACTACACCCGACGAATTCGTGATGAGCTGACATTTCCCGGAGACATCTCTG 261

QY      386  AlaValIleLeuGlnGluGlyGluAspGlyTyrTrpThrValGlnArGAnGlyGln 403
      |||||||
Db      260  GCGGTCTCTCTCGAGAGGGAGGATGCTGCTGTGACTGTGAGCGGACGACACAA 207

RESULT 15
BM933148 543 bp  mRNA  linear  EST 13-MAR-2002
LOCUS      BM933148
DEFINITION UT-M-BH3-brx-b-10-0-UT.k1 NIH_BMAP_M_S4 Mus musculus cDNA clone
VERSION     UT-M-BH3-brx-b-10-0-UT 5', mRNA sequence.
ACCESSION  BM933148  GI:19392300
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
REFERENCE  1 (bases 1 to 543)
AUTHORS   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
          Bonaldo,M.F., Lennon,G. and Soares,M.B.
          Normalization and subtraction: two approaches to facilitate gene
          discovery
          Genome Res. 6 (9), 791-806 (1996)
JOURNAL   97044477
MEDLINE   Contact: Chin, H
          National Institute of Mental Health
          6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
          20892-9643, USA
          Tel.: 301 443 1706
          Fax: 301 443 9890
          Email: mestr@mail.nih.gov
          cDNA library preparation: Dr. M. Bento Soares, University of Iowa
          cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
          DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
          Clone Distribution: Researchers may obtain clones from Research
          Genetics (www.resgen.com).
          Seq primer: M13 REVERSE.
FEATURES
source     Location/Qualifiers
1..543
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UT-M-BH3-brx-b-10-0-UT"
/clone_1lb="NIH_BMAP_M_S4"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The
NIH_BMAP_M_S4 library is a subtracted library of a series,

```

ultimately derived from a mixture of individually tagged normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH_BMAP_M_S4, NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1, NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library (NIH_BMAP_M_S4) was constructed as follows: PCR amplified cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with a pool of the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 libraries in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (liferotechnologies) to generate the NIH_BMAP_M_S4 library. This procedure has been previously described (Bonaldi, Lennon and Soares, Genome Research 6:791-806, 1996).

BASE COUNT 122 a 160 c 144 g 115 t 2 others
ORIGIN

Alignment Scores:

Pred. No.: 6.56e-58 Length: 543
Score: 69.00 Matches: 128
Percent Similarity: 96.97% Conservative: 0
Best Local Similarity: 96.97% Mismatches: 2
Query Match: 16.63% Indels: 4
DB: 14 Gaps: 0

US-09-068-377c-1 (1-415) x BM933148 (1-543)

```

QY 208 G|UHisArGThrThrCysGluAlaPheGlnLeuGlnIupheAsPArgLeuThrIleLeu 227
    |||||||
Db 3 GAGCACCAGGACTACCTGTGAGCGCTTCCAGTGCAGAGTTTGACCGCTCACCATCCTC 62

QY 228 ArgAsnAlaLeuTrpValHisCysAsnGlnLeuSerMetGlnCysValLysAspAspGlu 247
    |||||||
Db 63 CGCAATGCCCTGTGGGTGCACACTGTAACCACTCTCCAGTGTGTCAAGATGATGAG 122

QY 248 LeuTyrgluGluValArgLeuThrLeuGlnGlyCysAspValGluGlyAspIleAsnGly 267
    |||||||
Db 123 CTCTATGAGGAGAGCGCGCTGACCCCTTGAGGCGTGTGATGTGGAAGTGACATCAATGGC 182

QY 268 PheIleGlnSerLysSerThrGlyArgGlu-ProProAlaProValProTyrGlnAsnTy 287
    |||||||
Db 183 TTCATCCAGTCCAGAGACTGGAGATA-GCCCCAGCTCCGGTGCCTTATCAGACTA 241

QY 287 rTyAsPArgGluValThrProLeuIleGlySerProSerIle-GlnProSerCysGlyV 307
    |||||||
Db 242 CTATGACAGGAGAGTGAACCCACTGATGACAGCCCTAGCGT-CCACCCCTCCTGTGGTG 300

QY 307 alleLysArgPheSerGlyLeuLeuHisGlySerProLysThrThrProSerAlaProA 327
    |||||||
Db 301 TGAATAAGAGAGCTTCTGGGCTGCTACATGAAATCCCAAGACCAACACTTCTGCTCTG 360

QY 327 laAlaSerThrGluThrLeuThrProThrPro 337
    |||||||
Db 361 CTGCTTCCACAGAGACTGTGACTGCCACCCCT 392

```

Search completed: August 3, 2003, 16:34:06
Job time : 1862 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 3, 2003, 15:00:32 ; Search time 67 Seconds
(without alignments)
1899.565 Million cell updates/sec

Title: US-09-068-377C-1

Sequence: 1 MMAQLQFRDAFWCRDTAHT.....WTVERNGQGRGFPVPSYLEKL 415

Scoring table: OLIGO

Xgapop	60.0	Xgapext	60.0
Ygapop	60.0	Ygapext	60.0
Fgapop	6.0	Fgapext	7.0
Delop	6.0	Delext	7.0

Searched: 441362 seqs, 153338381 residues

Word size:

Total number of hits satisfying chosen parameters: 878600

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Listing first 45 summaries

Command line parameters:

```

MODEL=framed_p2n.model -DEV=fasta
-O=cnrn1.usvpo.spool/US09068377/runat.28072003.091404.15931/app.query.fasta.1.588
-DB=update Patents.NA -QFM=fastap -SUPPLX=ol1p2n.rm1 -MINMATCH=0.1 -LOOCL=0
-LOOPEX=0 -UNITS=bits -START=1 -END=1 -MATRIX=ol1lgo -TRANS=human40.cdi
-LIST=45 -DOCL1G=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=DOCL
-OUFMT=ptc -NORM=ecore -HEAPSIZ=500 -MINDEN=0 -MALEXN=2000000000
-USER=US09068377.ecgn.1.1.40.tunat.28072003.091404.15931 -NCPU=6
-NO_MMAP -LAREJORY -NMG_SCORES=0 -WAIT -DSBPLACK=100 -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREDS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

```

Database : Issued_patents_NA:*

```
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
2:
3: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6A_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/6B_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq: *
7: /cgn2_6/ptodata/1/ina/Backfiles1.seq:
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query	Match	Length	DB	ID	Description
No.							
1	415	100.0	2100	3	US-08-938-830-2		Sequence 2, App1
2	415	100.0	2100	3	US-09-020-222-2		Sequence 2, App1
3	31	7.5	1613	3	US-08-938-830-28		Sequence 28, App
4	31	7.5	1803	4	US-09-006-428A-18		Sequence 18, App
5	31	7.5	1858	4	US-09-006-428A-16		Sequence 16, App
6	10	2.4	457	3	US-08-938-830-42		Sequence 42, App
7	10	2.4	48	3	US-08-938-830-40		Sequence 40, App
8	10	2.4	48	3	US-08-938-830-41		Sequence 41, App
9	9	2.2	45	3	US-08-938-830-39		Sequence 39, App
10	8	1.9	37	3	US-08-938-830-16		Sequence 16, App
11	8	1.9	37	3	US-09-020-222-16		Sequence 16, App
12	8	1.9	42	3	US-08-938-830-30		Sequence 30, App

C 13	8	1.9	42	3	US-08-938-830-31	Sequence 31, Appl
C 14	8	1.9	42	3	US-08-938-830-32	Sequence 32, Appl
C 15	8	1.9	42	3	US-08-938-830-33	Sequence 33, Appl
C 16	8	1.9	42	3	US-08-938-830-34	Sequence 34, Appl
C 17	8	1.9	42	3	US-08-938-830-35	Sequence 35, Appl
C 18	8	1.9	42	3	US-08-938-830-37	Sequence 37, Appl
C 19	8	1.9	42	3	US-08-938-830-55	Sequence 55, Appl
C 20	8	1.9	646	4	US-08-998-416-77	Sequence 77, Appl
C 21	8	1.9	1545	4	US-08-759-579-1	Sequence 1, Appl
C 22	8	1.9	1845	1	US-08-014-531-1	Sequence 1, Appl
C 23	8	1.9	1845	1	US-08-262-338-1	Sequence 1, Appl
C 24	8	1.9	1845	1	US-08-460-114-1	Sequence 1, Appl
C 25	8	1.9	1845	4	US-09-347-878-37	Sequence 37, Appl
C 26	8	1.9	1845	4	US-09-133-966-1	Sequence 1, Appl
C 27	8	1.9	2865	2	US-08-749-169-2	Sequence 2, Appl
C 28	8	1.9	2865	2	US-09-130-032-2	Sequence 2, Appl
C 29	8	1.9	4335	3	US-09-058-489-19	Sequence 19, Appl
C 30	8	1.9	4425	2	US-08-749-169-1	Sequence 1, Appl
C 31	8	1.9	4425	2	US-09-130-032-1	Sequence 1, Appl
C 32	8	1.9	4931	3	US-09-058-489-20	Sequence 20, Appl
C 33	8	1.9	6476	3	US-09-058-489-21	Sequence 21, Appl
C 34	8	1.9	4403765	4	US-09-103-840A-2	Sequence 2, Appl
C 35	8	1.9	4411529	4	US-09-103-840A-1	Sequence 1, Appl
C 36	7	1.7	33	3	US-08-938-830-9	Sequence 9, Appl
C 37	7	1.7	33	3	US-09-020-222-9	Sequence 9, Appl
C 38	7	1.7	36	3	US-08-938-830-58	Sequence 58, Appl
C 39	7	1.7	40	3	US-08-938-830-38	Sequence 38, Appl
C 40	7	1.7	42	3	US-08-938-830-59	Sequence 59, Appl
C 41	7	1.7	71	4	US-08-952-793-371	Sequence 371, Appl
C 42	7	1.7	71	5	PCT-US96-09455A-371	Sequence 371, Appl
C 43	7	1.7	239	5	PCT-US93-08106-7	Sequence 7, Appl
C 44	7	1.7	239	5	PCT-US93-08106-7	Sequence 7, Appl
C 45	7	1.7	239	5	PCT-US94-00089-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1
 US-08-938-830-2
 Sequence 2, Application US/08938830
 Patent No. 6040437
 GENERAL INFORMATION:
 APPLICANT: Lasky, Laurence A.
 APPLICANT: Dombenko, Donald J.
 TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage
 TITLE OF INVENTION: Furrow-Associated Proteins (FSTPIS)
 NUMBER OF SEQUENCES: 73
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genentech, Inc.
 STREET: 1 DNA Way
 CITY: South San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94080
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: WinPatIn (Genentech)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/938,830
 FILING DATE:
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/798419
 FILING DATE: 07-FEB-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Dreger, Ginger R.
 REGISTRATION NUMBER: 33,055
 REFERENCE/DOCKET NUMBER: P1066P1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 650/225-3216
 TELEFAX: 650/952-9881

```

: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 2100 base pairs
:   TYPE: Nucleic Acid
:   STRANDEDNESS: Single
:   TOPOLOGY: Linear
: US-08-938-830-2
:
: Alignment Scores:
: Pred. No.: 0          Length: 2100
: Score: 415.00         Matches: 415
: Percent Similarity: 100.00%      Conservative: 0
: Best Local Similarity: 100.00%    Mismatches: 0
: Query Match: 100.00%             Indels: 0
: DB: 3                       Gaps: 0
:
US-09-068-377c-1 (1-415) x US-08-938-830-2 (1-2100)
QY      1 MetMetAlaGlnLeuGlnPheArgAspAlaPheTrpCysArgAspPheThrAlaHisThr 20
Db      682 ATGATGGCCCGCAGCTCCGAGTCCGATGCTTGTGTGACAGGAGCTTCACGGCCACACA 741
QY      21 G1TGTG1uValLeuLeuGlnArgLeuLeuAspGlyArgLysMetCysLysAspValGlu 40
Db      742 GGGATGATGAGGTGCTACTGACAGAGGCTCTGACGCGCAGGAAGATGTGCAAGGATGTGAG 801
QY      41 GluLeuLeuArgGlnArgAlaGlnAlaGluGluArgTyrGlyLysGluLeuValGlnIle 60
Db      802 GAGCTGTCTCAGACAGAGGCGCCAGCGGAGAGAGTACCGGAAGAGAGCTGTGTCAATT 861
QY      61 AlaArgLysAlaGlyGlnThrGluMetAsnSerLeuArgThrPheAspSerLeu 80
Db      862 GCACGCAAGCCTGCTGGCCAGACAGAGATGATTCCTTGAGGACCTCTTGACTCCCTG 921
QY      81 LysGlnGlnThrGluAsnValGlySerAlaHisIleGlnLeuAlaLeuArgGlu 100
Db      922 AAGCAGCAACACAGATGATGTGGCAGTGCACACATTCAGCTGGCCCTGCGCTGTGAG 981
QY      101 GluLeuArgSerLeuGlnPheArgGluArgGlnLysGlnArgLysLysLysTyrGlu 120
Db      982 GAGCTGGGAGCCTGGAGGAGTTCGCGAGAGACAGAAAGACGCGGAGGAAGTATGAG 1041
QY      121 AlaIleMetLysPargValGlnLysSerLysLeuSerLeuTyrLysLysThrMetGluSer 140
Db      1042 GCCATCATGACCGCTGCCAGAGAGCAAGTCTGCTTCAACAAGACCATGAGATGCC 1101
QY      141 LysLysAlaIleTyrAspGlnLysCysArgAspAlaAspAspAlaGlnGlnAlaPheGluArg 160
Db      1102 AAGAAAGCATATGACCAAGAGAGTCCAGGATGACAGATGATCTGTGAGCGCTTCGAGCGT 1161
QY      161 ValSerAlaAsnGlnLysGlnValGlnLysSerGlnAsnLysAlaLysGlnCys 180
Db      1162 GTGAGTGGCAATGGCCACCAAGCAAGTGAAGAAAGCCAGAAACCAAGCAGAGCTGC 1221
QY      181 LysGlnLysAlaThrGlnAlaGlnValTyrArgGlnAsnIleGlnLysGluArg 200
Db      1222 AAGAGACTGACCCCAAGAGGAGAAAGAGTGTACAGGCAAAATATCGAACAACATGGAGAGA 1281
QY      201 AlaArgThrGlnIleArgGlnGlnLysArgThrThrCysGlnAlaPheGlnLeuGlnGlu 220
Db      1282 GCCAGAGCCGAGTGGAGGAGGACACCGGACTTACCTGTGAGGCTTCCTCCAGTTCGAGGAG 1341
QY      221 PheAspArgLeuThrIleLeuArgAsnAlaLeuTrpValHisCysAsnGlnLeuSerMet 240
Db      1342 TTGTGACCGGCTCACCATTCCCGCAATGCCCTGTGGTGTGACGTAAACGAGCTCTCATG 1401
QY      241 GlnCysValLysAspAspGlnLeuTyrGlnGlnValArgLeuThrLeuGlnGlyCysAsp 260
Db      1402 CAGTGTGTCAAGAGATATGAGCTCTATGAGGAAGTCCGGCTGACCTTGAGGCGCTGTGAT 1461
QY      261 ValGlnGlyAspIleAsnGlnPheIleGlnSerLysSerThrLysArgGluProPheAla 280
Db      1462 GTGGAAAGGTGACATGAGCTTCAATCCAGTCCAAAGACACTGGCAGAGAGCCCGCAGCT 1521
```

```

QY      281 ProValProTyrGlnAsnTyrTyrAspArgGluValThrProLeuIleGlySerProSer 300
Db      1522 CCGGTGCTTATCAGAACTACTATGACAGGAGGTGACCCACATGATGGACGCCCTAGC 1581
QY      301 IleGlnProSerCysGlyValIleLysArgPheSerGlyLeuLeuHisGlySerProLys 320
Db      1582 ATCCAGCCCTCCTGCGGTGTGATTAAGAGGTCTCTGTGGCTGTGACTGAGAGTCCCAAG 1641
QY      321 ThrThrProSerAlaProAlaAlaSerThrGluThrLeuThrProThrProGluArgAsn 340
Db      1642 ACCACACCTTCTGCTCTGCTGCTTCCACAGAGACTGTGACTCCACCCCTGAGCGGAAAT 1701
QY      341 GluLeuValLysAlaSerIleGlnValGlnAlaThrGlnGlnLysAsnLeuAsnSerSerAla 360
Db      1702 GAGTTGTCTTACCATTCATTCAGAGTCCAGGCCAGCCAGGGAACCTTAATCTCATGAGCC 1761
QY      361 GlnAspTyrArgAlaLeuTyrAspTyrThrAlaGlnAsnSerAspGluLeuAspIleSer 380
Db      1762 CAGGACTACCGGGCAGCTCTACGACTACACTGACAGAGATTCGTGATGAGCTGGACATTTCC 1821
QY      381 AlaGlyAspIleLeuAlaValIleLeuGlnGlyGlnAspGlyTyrTrpThrValGluArg 400
Db      1822 GCGGAGACATCCCTGGGGTATCTCGGAAGGAGGATGCTGTGAGACTGTGAGACGG 1881
QY      401 AsnGlyGlnArgGlyPheValProGlySerTyrLeuGlnLysLeu 415
Db      1882 AACGGAACAAGTGGCTTGTTCCTCGGTGCTGACTTGAGAGAAGCTC 1926

RESULT 2
US-09-020-222-2
: Sequence 2, Application US/09020222
: Patent No. 611073
:
: GENERAL INFORMATION:
:   APPLICANT: Lasky, Laurence A.
:   TITLE OF INVENTION: Tyrosine phosphorylated Cleavage
:   TITLE OF INVENTION: Furrow-Associated Proteins (PSTPIPS)
:   NUMBER OF SEQUENCES: 27
:   CORRESPONDENCE ADDRESSES:
:     ADDRESSEE: Genentech, Inc.
:     STREET: 1 DNA Way
:     CITY: South San Francisco
:     STATE: California
:     COUNTRY: USA
:     ZIP: 94080
:
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: WinPatlin (Genentech)
:   CURRENT APPLICATION DATA:
:     APPLICATION NUMBER: US/09/020,222
:     FILING DATE: 06-Feb-1998
:
: CLASSIFICATION:
:   PRIOR APPLICATION DATA:
:     APPLICATION NUMBER: 08/798419
:     FILING DATE: 02/07/1997
:   ATTORNEY/AGENT INFORMATION:
:     NAME: Dreyer, Ginger R.
:     REGISTRATION NUMBER: 33,055
:     REFERENCE/DOCKET NUMBER: P1066r1
:   TELECOMMUNICATION INFORMATION:
:     TELEPHONE: 650/225-3216
:     TELEFAX: 650/952-9881
:
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 2100 base pairs
:   TYPE: Nucleic Acid
:   STRANDEDNESS: Single
:   TOPOLOGY: Linear
: US-09-020-222-2
:
: Alignment Scores:
```

```

Pred. No.: 0 Length: 2100
Score: 415.00 Matches: 415
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-09-068-377C-1 (1-415) x US-09-020-222-2 (1-2100)

QY 1 MetMetAlaGlnLeuGlnPheArGAspAlaPheTrpCysArGAspPheThrAlaHisThr 20
   |||||
Db 682 ATGATGCCCCAGCTGCGAGTCCGAGATGCTTGTGGCGAGGACTTCACGGCCACACA 741
   |||||

QY 21 GLTyrGlnValLeuLeuGlnArGLeuAspGlyArgLysMetCysLysAspValGlu 40
   |||||
Db 742 GGTATGAGGTGCTACTGCGAGAGCTGCTGGACGGACGAAAGTGTCAAGATGTGGAG 801
   |||||

QY 41 GlnLeuLeuArGlnArGlnAlaGlnGluArgTyrGlyLysGlnLeuValGlnIle 60
   |||||
Db 802 GAGCTGCTCAGACAGAGGGGCCAGGGGAGAGAGAGTACGGCAAGGAGCTGTGCAGATT 861
   |||||

QY 61 AlaArgLysAlaGlyGlyGlnThrGluMetAsnSerLeuArGThrSerPheAspSerLeu 80
   |||||
Db 862 GCACGCAAGGCTGTGCGCCAGACAGATGAATTCCCTGAGGACTCTTGTGACTCCCTG 921
   |||||

QY 81 LysGlnGlnThrGlnAsnValGlySerAlaHisIleGlnLeuAlaLeuArGLeu 100
   |||||
Db 922 AAGCAGCAACAGAGATGTGGCGAGTGCACACATCCAGCTGGCCCTGCGCGAG 981
   |||||

QY 101 GlnLeuArGSerLeuGlnGluPheArGLeuArgGlnLysGlnLeuArGlyLysTyrGlu 120
   |||||
Db 982 GAGCTCGGAGCTGTGGAGAGTCCGAGAGAGACAGAAAGCAGCAGGAGAGATAGAG 1041
   |||||

QY 121 AlaIleMetAspArgValGlnLysSerLysLeuSerLeuTyrLysLysThrMetGluSer 140
   |||||
Db 1042 GCCATATGAGACCGTGTCCAGAGACCAAGTTCTGCTACAGAAAGACCATGAGTCC 1101
   |||||

QY 141 LysLysAlaTyrAspGlnLysCysArgAspAlaAspAspAlaGlnGlnAlaPheGluArg 160
   |||||
Db 1102 AAGAGGCAATATGACCAAGAGTGCAGAGATGCAGATGATGTCTAGCGAGGCTTCGAGCGGT 1161
   |||||

QY 161 ValSerAlaAsnGlnHisGlnLysGlnValGlnLysSerGlnAsnLysAlaLysGlnCys 180
   |||||
Db 1162 GTGAGTGCCAATGCGCCAGCAGAGCAAGTGAAGAAAGCCAGAAAGCCAGAGTGC 1221
   |||||

QY 181 LysGlnSerAlaThrGlnAlaGlnArgValTyrArgGlnAsnIleGlnLeuGluArg 200
   |||||
Db 1222 AAGGAGTCAAGCCACAGAGCGAGAAAGTGTACAGGCCAAATATCGAACAACAGGAGAGA 1281
   |||||

QY 201 AlaArgThrGlnLysGlnGlnLysArgThrThrCysGlnAlaPheGlnLeuGlnGlu 220
   |||||
Db 1282 GCGAGAGACCGAGGAGGAGCAGAGAGCCGAGACTACCTGTGAGGCTTCCAGTTGCGAGGAG 1341
   |||||

QY 221 PheAspArgLeuThrIleLeuArGAsnAlaLeuTyrValHisCysAsnGlnLeuSerMet 240
   |||||
Db 1342 TTTGACCGGCTCAACATCTCCGCAATGCCCTGTGGGTGACGTAAACAGCTCTCCATG 1401
   |||||

QY 241 GlnCysValLysAspAspGlnLeuTyrGlnGlnValArgLeuThrLeuGlnGlyCysAsp 260
   |||||
Db 1402 CAGTGTGTCAAGATATAGCTCTATGAGGAAGTCCGCGTCACTTGAAGGCTGTGAT 1461
   |||||

QY 261 ValGlnGlyAspIleAsnGlnPheIleGlnSerLysSerThrGlyArgGluProPheAla 280
   |||||
Db 1462 GTGGAAGGTGACATCATGCTTCATCCAGTCCAAAGAGCAGTGGCAGAGAGCCCCAGCT 1521
   |||||

QY 281 ProValProTyrGlnAsnTyrTyrAspArgGluValThrProLeuIleGlySerProSer 300
   |||||
Db 1522 CCGGTGCTTATATCAAGTACTATGACAGGAGGTGACCCCACTGATTTGGCAGCCCTAGC 1581
   |||||

QY 301 IleGlnProSerCysGlyValIleLysArgPheSerGlyLeuLeuHisGlySerProLys 320
   |||||
Db 1582 ATCCAGCCCTCCGCGGTGTGATTAAGAGATTCTCTGGGGCTGTACATGGAAGTCCCAAG 1641
   |||||

```

```

QY 321 ThrThrProSerAlaProAlaAlaSerThrGlnThrLeuThrProThrProGluArgAsn 340
   |||||
Db 1642 ACCACACCTTGTGCTCTGCTGCTGCCACAGACTGTGACTCCACCCCTGAGCGGAGAT 1701
   |||||

QY 341 GlnLeuValTyrAlaSerIleGlnValGlnAlaThrGlnGlnLysLeuAsnSerSerAla 360
   |||||
Db 1702 GAGTTGGTCTACGATCATCATGAGTGCAGGAGCCAGGAGAAACCTTAATCATTCAGCC 1761
   |||||

QY 361 GlnAspTyrAlaGlnAlaLeuTyrAspTyrThrAlaGlnAsnSerAspGlnLeuAspIleSer 380
   |||||
Db 1762 CAGGACTACCGGCGCAGCTACGACTACAGTACAGCAAGAAATTCATGATGAGCTGACATTTC 1821
   |||||

QY 381 AlaGlyAspIleLeuAlaValIleLeuGlnGlyLysAspGlyTyrPheThrValGluArg 400
   |||||
Db 1822 GCGGAGACATCTCCGCGGTATCTCTGAAAGGAGGAGATGCTGTGAGTGTGAGCGG 1881
   |||||

QY 401 AsnGlyGlnArGlyPheValProGlySerTyrLeuGlnLysLeu 415
   |||||
Db 1882 AAGGACACAGCTGCTTGTTCCTCGGGTGTGACTTGAGAAAGCTC 1926
   |||||

RESULT 3
US-08-938-830-28
: Sequence 28, Application US/08938830
: Patent No. 6040437
: GENERAL INFORMATION:
: APPLICANT: Lasky, Laurence A.
: TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage
: NUMBER OF INVENTION: Furrow-Associated Proteins (PSRIPs)
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 1 DNA Way
: CITY: South San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94080
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: WinPatIn (Genentech)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/938, 830
: FILING DATE:
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/798419
: FILING DATE: 07-FEB-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Dreger, Ginger R.
: REGISTRATION NUMBER: 33,055
: REFERENCE/DOCKET NUMBER: P1066P1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650/225-3216
: TELEFAX: 650/952-9881
: INFORMATION FOR SEQ ID NO: 28:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1613 base pairs
: TYPE: Nucleic Acid
: STRANDEDNESS: Single
: TOPOLOGY: Linear
: US-08-938-830-28

Alignment Scores:
Pred. No.: 1,54e-21 Length: 1613
Score: 31.00 Matches: 60
Percent Similarity: 96.77% Conservative: 0
Best Local Similarity: 96.77% Mismatches: 1
Query Match: 7.47% Indels: 2
DB: 3 Gaps: 0

US-09-068-377C-1 (1-415) x US-08-938-830-28 (1-1613)

```


APPLICATION NUMBER: 08/798419
FILING DATE: 07-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: P1066P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-3216
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 47 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
US-08-938-830-42

Alignment Scores:
Pred. No.: 0.0438 Length: 47
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.41% Indels: 0
Gaps: 0
DB: 3

US-09-068-377c-1 (1-415) x US-08-938-830-42 (1-47)

OY 91 HistHleGlInleuAlaLeuAlaLeuArgIu 100
Db 47 CACATCCAGCTGCGCTGCGCTGCGGAG 18

RESULT 7
US-08-938-830-40/c
Sequence 40, Application US/08938830
Patent No. 6040437
GENERAL INFORMATION:
APPLICANT: Lasky, Laurence A.
APPLICANT: Dowbenko, Donald J.
TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage
TITLE OF INVENTION: Furrow-Associated Proteins (PSTPIPs)
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938,830
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/798419
FILING DATE: 07-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: P1066P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-3216
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 48 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear

US-08-938-830-40

Alignment Scores:
Pred. No.: 0.0446 Length: 48
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.41% Indels: 0
Gaps: 0
DB: 3

US-09-068-377c-1 (1-415) x US-08-938-830-40 (1-48)

OY 255 ThrleuGluglGlyCysAspValGluglYasp 264
Db 47 ACCCTTGAGGCGTGTGATGTGGAAGCTGAC 18

RESULT 8
US-08-938-830-41/c
Sequence 41, Application US/08938830
Patent No. 6040437
GENERAL INFORMATION:
APPLICANT: Lasky, Laurence A.
APPLICANT: Dowbenko, Donald J.
TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage
TITLE OF INVENTION: Furrow-Associated Proteins (PSTPIPs)
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938,830
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/798419
FILING DATE: 07-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: P1066P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-3216
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 48 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
US-08-938-830-41

Alignment Scores:
Pred. No.: 0.0446 Length: 48
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.41% Indels: 0
Gaps: 0
DB: 3

US-09-068-377c-1 (1-415) x US-08-938-830-41 (1-48)

OY 230 AlaLeuTrpValHisCysAsnGlnLeuSer 239
Db 47 GCCCTGTGGCTGACACTGTAACCACTCTCC 18

RESULT 9
US-08-938-830-39/c
; Sequence 39, Application US/08938830
; Patent No. 6040437
; GENERAL INFORMATION:
; APPLICANT: Lasky, Laurence A.
; APPLICANT: Dowbenko, Donald J.
; TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage
; TITLE OF INVENTION: Furrow-Associated Proteins (PSTPIPs)
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Winpallin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,830
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/798419
; FILING DATE: 07-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: P1066P1
; TELEPHONE: 650/225-3216
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; US-08-938-830-39

Alignment Scores:
Pred. No.: 0 414 Length: 45
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.17% Indels: 0
Gaps: 0
DB: 3 0

US-09-068-377C-1 (1-415) x US-08-938-830-39 (1-45)

QY 281 ProValProTyrGlnAsnTyrTyrAsp 289
Db 44 CCGGTCCTTATCAGACTACTATGAC 18

RESULT 10
US-08-938-830-16
; Sequence 16, Application US/08938830
; Patent No. 6040437
; GENERAL INFORMATION:
; APPLICANT: Lasky, Laurence A.
; APPLICANT: Dowbenko, Donald J.
; TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage
; TITLE OF INVENTION: Furrow-Associated Proteins (PSTPIPs)
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco

STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpallin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938,830
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/798419
FILING DATE: 07-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: P1066P1
TELEPHONE: 650/225-3216
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
US-08-938-830-16

Alignment Scores:
Pred. No.: 3 41 Length: 37
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.93% Indels: 0
Gaps: 0
DB: 3 0

US-09-068-377C-1 (1-415) x US-08-938-830-16 (1-37)

QY 365 AlaLeuTyrAspTyrThrAlaGln 372
Db 14 GCACTCTACGACTACTGACAG 37

RESULT 11
US-09-020-222-16
; Sequence 16, Application US/09020222
; Patent No. 611073
; GENERAL INFORMATION:
; APPLICANT: Lasky, Laurence A.
; TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage
; TITLE OF INVENTION: Furrow-Associated Proteins (PSTPIPs)
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Winpallin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/020,222
; FILING DATE: 06-Feb-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/798419
; FILING DATE: 02/07/1997
; ATTORNEY/AGENT INFORMATION:

NAME: Dregler, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: P1066r1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-3216
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear

US-09-020-222-16

Alignment Scores:

Pred. No.:	3 41	Length:	37
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.93%	Indels:	0
DB:	3	Gaps:	0

US-09-068-377c-1 (1-415) x US-09-020-222-16 (1-37)

OY 365 AlaLeuTyAspTyrThrAlaGln 372
|||||
DB 14 GCACCTACGACTACACGCGACAG 37

RESULT 12

US-08-938-830-30
Sequence 30, Application US/08938830
Patent No. 6040437

GENERAL INFORMATION:
APPLICANT: Lasky, Laurence A.
APPLICANT: Dowbenko, Donald J.
TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage
TITLE OF INVENTION: Furorow-Associated Proteins (PSTPIPS)
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESS: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938,830
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/798419
FILING DATE: 07-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Dregler, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: P1066r1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-3216
TELEFAX: 650/952-9881

INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 42 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear

US-08-938-830-30

Alignment Scores:

Pred. No.:	3 83	Length:	42
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.93%	Indels:	0
DB:	3	Gaps:	0

US-09-068-377c-1 (1-415) x US-08-938-830-30 (1-42)

OY 26 LeuGlnArgLeuLeuAspGlyArg 33
|||||
DB 19 CTGACAGGCGCTGTCGACGCGACAG 42

RESULT 13

US-08-938-830-31
Sequence 31, Application US/08938830
Patent No. 6040437

GENERAL INFORMATION:
APPLICANT: Lasky, Laurence A.
APPLICANT: Dowbenko, Donald J.
TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage
TITLE OF INVENTION: Furorow-Associated Proteins (PSTPIPS)
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESS: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938,830
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/798419
FILING DATE: 07-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Dregler, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: P1066r1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-3216
TELEFAX: 650/952-9881

INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 42 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear

US-08-938-830-31

Pred. No.:	3 83	Length:	42
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.93%	Indels:	0
DB:	3	Gaps:	0

US-09-068-377c-1 (1-415) x US-08-938-830-31 (1-42)

OY 51 GluArgTyrGlyGluLeuVal 58
|||||
DB 19 GAGAGGTACGGAGGAGCTGGTG 42

RESULT 14

US-08-938-830-32

Sequence 32, Application US/08938830
Patent No. 6040437
GENERAL INFORMATION:
APPLICANT: Lasky, Laurence A.
APPLICANT: Dowbenko, Donald J.
TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage
TITLE OF INVENTION: Furrow-Associated Proteins (PSTPIPs)
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938,830
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/798419
FILING DATE: 07-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: P1066P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-3216
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 42 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
US-08-938-830-32
Alignment Scores:
Pred. No.: 3.83 Length: 42
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.93% Indels: 0
DB: 3 Gaps: 0
US-09-068-377C-1 (1-415) x US-08-938-830-32 (1-42)
QY 76 SerPheAspSerLeuylsGlnIn 83
Db 19 TCCTTTGACTCCCTGAAGCAGCA 42
RESULT 15
US-08-938-830-33
Sequence 33, Application US/08938830
Patent No. 6040437
GENERAL INFORMATION:
APPLICANT: Lasky, Laurence A.
APPLICANT: Dowbenko, Donald J.
TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage
TITLE OF INVENTION: Furrow-Associated Proteins (PSTPIPs)
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938,830
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/798419
FILING DATE: 07-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: P1066P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-3216
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 42 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
US-08-938-830-33
Alignment Scores:
Pred. No.: 3.83 Length: 42
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.93% Indels: 0
DB: 3 Gaps: 0
US-09-068-377C-1 (1-415) x US-08-938-830-33 (1-42)
QY 101 GluLeuArgSerLeuGlnGluPhe 108
Db 19 GAGCTCGGAGGCTCGAGGAGTTTC 42

Search completed: August 3, 2003, 17:23:51
Job time : 82 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 3, 2003, 16:11:38 : Search time 305 Seconds
(without alignments)
2807.034 Million cell updates/sec

Title: US-09-068-377c-1
Perfect score: 415
Sequence: 1 MAAQLQFRDAFWCRDEFTAH.....WTVERNGRGVFGSYLEKL 415

Scoring table:
OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1439767 seqs, 1031500376 residues

Word size: 1

Total number of hits satisfying chosen parameters: 2875925

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_SPOOL/US09068377/runat_28072003_091406_16028/app_query.fasta_1.583
-DB=Published_Applications_NA -QPMT=fastlap -SUFFIX=olip2n.rnpb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=quality -THR MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09068377 -ECGN_1_1_221=runat_28072003_091406_16028
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published Applications_NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.1.*
- 11: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.2.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.3.*
- 13: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	31	7.5	1428	15	US-10-067-076-18 Sequence 18, Appl

2	31	7.5	1428	15	US-10-067-076-19	Sequence 19, Appl
3	31	7.5	1428	15	US-10-067-076-28	Sequence 28, Appl
4	30	7.2	1428	15	US-10-067-076-21	Sequence 21, Appl
5	29	7.0	1428	15	US-09-918-995-31326	Sequence 31326, A
6	29	7.0	488	11	US-09-867-701-10401	Sequence 10401, A
7	25	3.6	450	11	US-09-783-590-55775	Sequence 55777, Ap
8	10	2.4	523	15	US-10-027-632-174755	Sequence 174755,
9	10	2.4	776	15	US-10-027-632-162764	Sequence 162764,
10	9	2.2	616	15	US-10-027-632-99782	Sequence 99782, A
11	9	2.2	36221	12	US-09-954-556-29	Sequence 29, Appl
12	8	1.9	367	12	US-09-803-719-479	Sequence 479, Appl
13	8	1.9	447	11	US-09-833-381-1072	Sequence 1072, Ap
14	8	1.9	447	15	US-10-027-632-339709	Sequence 339709, A
15	8	1.9	447	15	US-10-027-632-294775	Sequence 294775,
16	8	1.9	463	15	US-10-027-632-82179	Sequence 82179, A
17	8	1.9	613	15	US-10-027-632-106048	Sequence 106048, A
18	8	1.9	626	15	US-10-027-632-20152	Sequence 20152, A
19	8	1.9	632	15	US-10-027-632-225122	Sequence 225122,
20	8	1.9	638	15	US-10-027-632-211381	Sequence 211381,
21	8	1.9	782	15	US-10-106-698-1282	Sequence 1282, Ap
22	8	1.9	822	15	US-10-027-632-156460	Sequence 156460,
23	8	1.9	948	11	US-09-738-626-508	Sequence 508, App
24	8	1.9	1220	10	US-09-770-445-38	Sequence 38, Appl
25	8	1.9	1275	11	US-09-738-626-516	Sequence 516, App
26	8	1.9	1530	15	US-10-142-231-54	Sequence 54, Appl
27	8	1.9	1803	11	US-09-822-830X-306	Sequence 306, App
28	8	1.9	2811	15	US-10-267-255-28	Sequence 28, Appl
29	8	1.9	2847	15	US-10-214-932-39	Sequence 39, Appl
30	8	1.9	3441	10	US-09-866-028-6	Sequence 6, Appl
31	8	1.9	3441	11	US-09-944-457-6	Sequence 6, Appl
32	8	1.9	3441	11	US-09-944-457-6	Sequence 6, Appl
33	8	1.9	3441	11	US-09-944-457-6	Sequence 6, Appl
34	8	1.9	3441	11	US-09-945-015-6	Sequence 6, Appl
35	8	1.9	3441	11	US-09-944-396-6	Sequence 6, Appl
36	8	1.9	3441	11	US-09-944-097-6	Sequence 6, Appl
37	8	1.9	3441	11	US-09-944-432-6	Sequence 6, Appl
38	8	1.9	3441	11	US-09-944-432-6	Sequence 6, Appl
39	8	1.9	3441	11	US-09-944-654-6	Sequence 6, Appl
40	8	1.9	3441	11	US-09-943-8514-6	Sequence 6, Appl
41	8	1.9	3441	11	US-09-944-413-6	Sequence 6, Appl
42	8	1.9	3441	11	US-09-944-403-6	Sequence 6, Appl
43	8	1.9	3441	11	US-09-944-896-6	Sequence 6, Appl
44	8	1.9	3441	11	US-09-944-944-6	Sequence 6, Appl
45	8	1.9	3441	11	US-09-944-929-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-10-067-076-18
Sequence 18, Application US/10067076
Publication No. US2003010440A1
GENERAL INFORMATION:
APPLICANT: Wise, Carol A.
TITLE OF INVENTION: Genetic Markers for Autoimmune Disorder
FILE REFERENCE: TEX871/4-006US/36000
CURRENT APPLICATION NUMBER: US/10/067, 076
CURRENT FILING DATE: 2002-02-04
PRIOR APPLICATION NUMBER: 60/287, 893
PRIOR FILING DATE: 2001-05-01
PRIOR APPLICATION NUMBER: 09/710, 693
PRIOR FILING DATE: 2000-11-08
NUMBER OF SEQ ID NOS: 28
SOFTWARE: Patent version 3.2
SEQ ID NO 18
LENGTH: 1428
TYPE: DNA
ORGANISM: Human Nucleic Acid
US-10-067-076-18
Alignment Scores: 5.82e-23 Length: 1428
Pred. No.: 31.00 Matches: 60
Score:

Percent Similarity: 96.77% Conservative: 0
Best Local Similarity: 96.77% Mismatches: 1
Query Match: 7.47% Indels: 2
DB: 15 Gaps: 0

US-09-068-377c-1 (1-415) x US-10-067-076-18 (1-1428)

QY 9 Aspalaphetrpcysarqasphethralahisthrglytyrgluvalleuleuglnarg 28
DB 25 GATGCCCTTTGGTGGCAGGAGCTTCACAGCCACAGGGGCTACAGAGTGTCTGCACGG 84
QY 29 Leuleuaspcllyarqlysmecyslyaspval-gluqluleuleuargclnargalagl 48
DB 85 CTCTCGATGCGCAGGAAGATGTGCAAGACAT-GGAGAGCTACTGTAGCGCAGAGGCCCA 143
QY 48 nalagluqluarqlytyrglysgluvalglnllealarglysalaglyglnth 68
DB 144 GGCGGAGAGCGGTACGGGAAGAGCTGTGCAGATGCGCAGGAAGCGAGGTGCCAGAC 203
QY 68 rglu 69
DB 204 GGAG 207

RESULT 2
US-10-067-076-19

; Sequence 19, Application US/10067076
; Publication No. US20030104404A1

; GENERAL INFORMATION:

; APPLICANT: wise, Carol A.

; TITLE OF INVENTION: Genetic Markers for Autoimmune Disorder

; FILE REFERENCE: TEX871/4-006US/36000

; CURRENT APPLICATION NUMBER: US/10/067,076

; CURRENT FILING DATE: 2002-02-04

; PRIOR APPLICATION NUMBER: 60/287,893

; PRIOR FILING DATE: 2001-05-01

; PRIOR APPLICATION NUMBER: 09/710,693

; PRIOR FILING DATE: 2000-11-08

; NUMBER OF SEQ ID NOS: 28

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 19

; LENGTH: 1428

; TYPE: DNA

; ORGANISM: Human Nucleic Acid

US-10-067-076-19

Alignment Scores:

Pred. No.: 5.82e-23

Score: 31.00

Percent Similarity: 96.77%

Best Local Similarity: 96.77%

Query Match: 7.47%

DB: 15

US-09-068-377c-1 (1-415) x US-10-067-076-19 (1-1428)

QY 9 Aspalaphetrpcysarqasphethralahisthrglytyrgluvalleuleuglnarg 28
DB 25 GATGCCCTTTGGTGGCAGGAGCTTCACAGCCACAGGGGCTACAGAGTGTCTGCACGG 84
QY 29 Leuleuaspcllyarqlysmecyslyaspval-gluqluleuleuargclnargalagl 48
DB 85 CTCTCGATGCGCAGGAAGATGTGCAAGACAT-GGAGAGCTACTGTAGCGCAGAGGCCCA 143
QY 48 nalagluqluarqlytyrglysgluvalglnllealarglysalaglyglnth 68
DB 144 GGCGGAGAGCGGTACGGGAAGAGCTGTGCAGATGCGCAGGAAGCGAGGTGCCAGAC 203
QY 68 rglu 69
DB 204 GGAG 207

RESULT 3
US-10-067-076-28

; Sequence 28, Application US/10067076
; Publication No. US20030104404A1

; GENERAL INFORMATION:

; APPLICANT: wise, Carol A.

; TITLE OF INVENTION: Genetic Markers for Autoimmune Disorder

; FILE REFERENCE: TEX871/4-006US/36000

; CURRENT APPLICATION NUMBER: US/10/067,076

; CURRENT FILING DATE: 2002-02-04

; PRIOR APPLICATION NUMBER: 60/287,893

; PRIOR FILING DATE: 2001-05-01

; PRIOR APPLICATION NUMBER: 09/710,693

; PRIOR FILING DATE: 2000-11-08

; NUMBER OF SEQ ID NOS: 28

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 28

; LENGTH: 1658

; TYPE: DNA

; ORGANISM: Human Nucleic Acid

US-10-067-076-28

Alignment Scores:

Pred. No.: 6.65e-23

Score: 31.00

Percent Similarity: 96.77%

Best Local Similarity: 96.77%

Query Match: 7.47%

DB: 15

US-09-068-377c-1 (1-415) x US-10-067-076-28 (1-1658)

QY 9 Aspalaphetrpcysarqasphethralahisthrglytyrgluvalleuleuglnarg 28
DB 241 GATGCCCTTTGGTGGCAGGAGCTTCACAGCCACAGGGGCTACAGAGTGTCTGCACGG 300
QY 29 Leuleuaspcllyarqlysmecyslyaspval-gluqluleuleuargclnargalagl 48
DB 301 CTCTCGATGCGCAGGAAGATGTGCAAGACAT-GGAGAGCTACTGTAGCGCAGAGGCCCA 359
QY 48 nalagluqluarqlytyrglysgluvalglnllealarglysalaglyglnth 68
DB 360 GGCGGAGAGCGGTACGGGAAGAGCTGTGCAGATGCGCAGGAAGCGAGGTGCCAGAC 419
QY 68 rglu 69
DB 420 GGAG 423

RESULT 4
US-10-067-076-21

; Sequence 21, Application US/10067076
; Publication No. US20030104404A1

; GENERAL INFORMATION:

; APPLICANT: wise, Carol A.

; TITLE OF INVENTION: Genetic Markers for Autoimmune Disorder

; FILE REFERENCE: TEX871/4-006US/36000

; CURRENT APPLICATION NUMBER: US/10/067,076

; CURRENT FILING DATE: 2002-02-04

; PRIOR APPLICATION NUMBER: 60/287,893

; PRIOR FILING DATE: 2001-05-01

; PRIOR APPLICATION NUMBER: 09/710,693

; PRIOR FILING DATE: 2000-11-08

; NUMBER OF SEQ ID NOS: 28

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 21

; LENGTH: 1428

; TYPE: DNA

; ORGANISM: Human Nucleic Acid

US-10-067-076-21

Alignment Scores:

Pred. No.: 6.67e-22

Score: 30.00

Percent Similarity: 96.77%

Best Local Similarity: 96.77%

Length: 1428
Matches: 60
Conservative: 0
Mismatches: 1

```
Query Match: 7.23% Indels: 2
DB: 15 Gaps: 0
US-09-068-377C-1 (1-415) x US-10-067-076-21 (1-1428)
QY 9 AspalAphetRPySaRgsaPhetRraIaHIsThGlyTyRgluValleuLeuGlnarg 28
   |||||||
DB 25 GATGCTTTGGGCGAGGGAGCTTCACAGCCACAGGGGCTAGAGGTGCTGCGACGG 84
QY 29 LeuLeuAspGlyArGlysmetCysLysAspVal-GluGluLeuLeuArgGlnArgAlaG 48
   |||||||
DB 85 CTTGTCGATGAGGAGGAGATGTGCAAGACAT-GGAGAGCTACTGAGCGAGGGGCCA 143
QY 48 naIaGluGluArGTYRglYLySgluLeuValGlnIleAlaArgLysAlaGlyGlnTh 68
   |||||||
DB 144 GCGGAGGAGGCGGTACGGAGAGAGCTGTGTCAGATCGCAGGAGGCGAGTGGCCAGAC 203
QY 68 rGlu 69
   ||||
DB 204 GGAG 207

RESULT 5
US-09-918-995-31326
; Sequence 31326, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: HySeq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 31326
; LENGTH: 473
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; LOCATION: (1)...(473)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-31326

Alignment Scores:
Pred. No.: 2,83e-21 Length: 473
Score: 29.00 Matches: 29
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.99% Indels: 0
DB: 12 Gaps: 0
US-09-068-377C-1 (1-415) x US-09-918-995-31326 (1-473)
QY 387 ValIleLeuGluGlyLysAspGlyTTPTrThValGlnArgAsnGlyGlnArgGlyPhe 406
   |||||||
DB 373 GTGATCTCGAAGGGAGGATGCTGTGTGACTGTGAGAGGAGCGGCGCTGGCTTC 432
QY 407 ValProGlySerTYRleuGluLysLeu 415
   |||||||
DB 433 GTCCCTGTTCTACTCTGAGAAAGCTT 459

RESULT 6
US-09-867-701-10401/C
; Sequence 10401, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
```

```
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10401
; LENGTH: 488
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(488)
; OTHER INFORMATION: n = A,T,C or G
US-09-867-701-10401

Alignment Scores:
Pred. No.: 2.91e-21 Length: 488
Score: 29.00 Matches: 29
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.99% Indels: 0
DB: 11 Gaps: 0
US-09-068-377C-1 (1-415) x US-09-867-701-10401 (1-488)
QY 387 ValIleLeuGluGlyLysAspGlyTTPTrThValGlnArgAsnGlyGlnArgGlyPhe 406
   |||||||
DB 270 GTGATCTCGAAGGGAGGAGATGCTGTGTGACTGTGAGAGGAGCGGCGCTGGCTTC 211
QY 407 ValProGlySerTYRleuGluLysLeu 415
   |||||||
DB 210 GTCCCTGTTCTACTCTGAGAAAGCTT 184

RESULT 7
US-09-783-590-5577
; Sequence 5577, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haselding, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5577
; LENGTH: 480
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (116)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (211)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (293)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (295)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (316)
```

```
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (317)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (326)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (328)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (329)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (331)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (338)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (357)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (359)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (366)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (367)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (371)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (375)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (389)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (394)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (395)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (403)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (409)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (415)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (418)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (419)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (424)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (431)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (443)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (468)
OTHER INFORMATION: n equals a,t,g, or c
```

```
NAME/KEY: misc feature
LOCATION: (471)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (478)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (479)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (487)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (489)
OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-5577

Alignment Scores:
Pred. No.: 1,99e-06 Length: 490
Score: 15.00 Matches: 15
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.61% Indels: 0
DB: 11 Gaps: 0

US-09-068-377c-1 (1-415) x US-09-783-590-5577 (1-490)
OY 24 ValdeuleuglnArgleuleuAspGlyArgLysMetCysLysASP 38
Db 90 GTGCTGCTGCAGCGCGCTTCTGGATGAGNAGAGATGTGCAGAACAC 134

RESULT 8
US-10-027-632-174755
Sequence 174755, Application us/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 174755
LENGTH: 523
TYPE: DNA
ORGANISM: Human
US-10-027-632-174755

Alignment Scores:
Pred. No.: 0.419 Length: 523
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.41% Indels: 0
DB: 15 Gaps: 0

US-09-068-377c-1 (1-415) x US-10-027-632-174755 (1-523)
```

```
OY 92 IIEGlnleuAlaleuAlaleuArgGlu 101
|||||
DB 436 ATACAACTGCTTACCTCAGAGAA 465

RESULT 9
US-10-027-632-162764/c
; Sequence 162764, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 162764
; LENGTH: 776
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-162764

Alignment Scores:
Pred. No.: 0.598 Length: 776
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.41% Indels: 0
DB: 15 Gaps: 0

US-09-068-377c-1 (1-415) x US-10-027-632-162764 (1-776)

OY 92 IIEGlnleuAlaleuAlaleuArgGlu 101
|||||
DB 87 ATACAACTGCTTACCTCAGAGAA 58

RESULT 10
US-10-027-632-99782/c
; Sequence 99782, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
```

```
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 99782
; LENGTH: 616
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-99782

Alignment Scores:
Pred. No.: 5.57 Length: 616
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.17% Indels: 0
DB: 15 Gaps: 0

US-09-068-377c-1 (1-415) x US-10-027-632-99782 (1-616)

OY 22 TyrgluValleuengluArgleu 30
|||||
DB 362 TATCAAGTCTACTACAGAGCTGTGTTA 336

RESULT 11
US-09-954-556-29
; Sequence 29, Application US/09954556
; Publication NO. US20030078219A1
; GENERAL INFORMATION:
; APPLICANT: Brett P. Monia
; APPLICANT: Susan M. Freier
; TITLE OF INVENTION: ANTISENSE MODULATION OF FIBROBLAST GROWTH FACTOR RECEPTOR 2 EX
; FILE REFERENCE: RRS-0250
; CURRENT APPLICATION NUMBER: US/09/954,556
; CURRENT FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 108
; SEQ ID NO 29
; LENGTH: 36221
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(36221)
; OTHER INFORMATION: n = A,T,C or G
US-09-954-556-29

Alignment Scores:
Pred. No.: 218 Length: 36221
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.17% Indels: 0
DB: 12 Gaps: 0

US-09-068-377c-1 (1-415) x US-09-954-556-29 (1-36221)

OY 273 SerThrGluArgGluProAlaPro 261
|||||
DB 6826 TCTACTGTAGTGAACCCCGGCCCA 6652

RESULT 12
US-09-803-719-479/c
; Sequence 479, Application US/09803719
; Publication NO. US20030044783A1
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Escobedo, Jaime
; APPLICANT: Innis, Michael A.
; APPLICANT: Garcia, Pablo Dominguez
; APPLICANT: Suduth-Klinger, Julie
; APPLICANT: Reinhard, Christoph
; APPLICANT: Giese, Klaus
; APPLICANT: Randazzo, Filippo
```

APPLICANT: Kennedy, Giulia C.
APPLICANT: Pot, David
APPLICANT: Kassam, Altaf
APPLICANT: Lamson, George
APPLICANT: Drmanac, Radoje
APPLICANT: Crkvenjakov, Radomir
APPLICANT: Dickson, Mark
APPLICANT: Drmanac, Snezana
APPLICANT: Labat, Ivan
APPLICANT: Leshkowitz, Dena
APPLICANT: Kita, David
APPLICANT: Garcia, Veronica
APPLICANT: Jones, Lee William
APPLICANT: Stache-Crain, Birgit
TITLE OF INVENTION: Human Genes and Gene Products
FILE REFERENCE: 1624.002
CURRENT APPLICATION NUMBER: US/09/803,719
CURRENT FILING DATE: 2001-03-09
PRIOR APPLICATION NUMBER: 60/188,609
PRIOR FILING DATE: 2000-03-09
NUMBER OF SEQ ID NOS: 2396
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 479
LENGTH: 367
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)---(367)
OTHER INFORMATION: n - A,T,C or G
US-09-803-719-479

Alignment Scores:
Pred. No.: 40.1 Length: 367
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.93% Indels: 0
DB: 12 Gaps: 0

US-09-068-377c-1 (1-415) x US-09-803-719-479 (1-367)

OY 332 ThrlcuthrProthrProgluarg 339
|||||
DB 327 ACCCTCACTCCACACCTGAGAGA 304

RESULT 13
US-09-833-381-1072
Sequence 1072, Application US/09833381
Patent No. US20020132090A1
GENERAL INFORMATION:
APPLICANT: Robison, Keith E.
TITLE OF INVENTION: No. US20020132090A1el Nucleic Acid and Protein Homologs
FILE REFERENCE: 5800-119
CURRENT APPLICATION NUMBER: US/09/833,381
CURRENT FILING DATE: 2001-04-11
PRIOR APPLICATION NUMBER: 09/516,448
PRIOR FILING DATE: 2000-02-29
NUMBER OF SEQ ID NOS: 2050
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1072
LENGTH: 447
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)---(447)
OTHER INFORMATION: n - A,T,C or G
US-09-833-381-1072

Alignment Scores:
Pred. No.: 47.8 Length: 447
Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.93% Indels: 0
DB: 11 Gaps: 0

US-09-068-377c-1 (1-415) x US-09-833-381-1072 (1-447)

OY 278 PropoalProvalProTyrGln 285
|||||
DB 262 CTCCTCAGCCCTGTCCATATCAA 285

RESULT 14
US-10-027-632-39709
Sequence 39709, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 39709
LENGTH: 447
TYPE: DNA
ORGANISM: Human
US-10-027-632-39709

Alignment Scores:
Pred. No.: 47.8 Length: 447
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.93% Indels: 0
DB: 15 Gaps: 0

US-09-068-377c-1 (1-415) x US-10-027-632-39709 (1-447)

OY 25 LeuLeuGlnArgLeuLeuaspGly 32
|||||
DB 96 CTCCTCAGAGACTGTGATGCA 119

RESULT 15
US-10-027-632-294775
Sequence 294775, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218
 ; PRIOR FILING DATE: 2000-02-24
 ; PRIOR APPLICATION NUMBER: US 60/167,363
 ; PRIOR FILING DATE: 1998-11-23
 ; PRIOR APPLICATION NUMBER: US 60/156,358
 ; PRIOR FILING DATE: 1999-09-28
 ; PRIOR APPLICATION NUMBER: US 60/146,002
 ; PRIOR FILING DATE: 1999-08-09
 ; NUMBER OF SEQ ID NOS: 325720
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO: 294775
 ; LENGTH: 447
 ; TYPE: DNA
 ; ORGANISM: Human
 US-10-027-632-294775

Alignment Scores:
 Pred. No.: 47.8 Length: 447
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.93% Indels: 0
 DB: 15 Gaps: 0

US-09-068-377C-1 (1-415) x US-10-027-632-294775 (1-447)

QY 25 LeuLeuGlnArgLeuLeuAspGly 32
 ||||||||||||||||||||
 Db 96 CTGCTGCAGAGACTGTGGATGGA 119

Search completed: August 3, 2003, 17:46:29
 Job time : 326 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2003, 09:15:22 ; Search time 39 Seconds
(without alignments)
2745.944 Million cell updates/sec

Title: US-09-068-377c-1

Perfect score: 2152
Sequence: 1 MAOQOFDAFCRCDFTAHT.....WTVNRNGRGFVPGSYLEKL 415

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_23:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.podent:*
12: sp.virus:*
13: sp.vertibrate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriap:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2152	100.0	415	11	P97814
2	1913.5	88.9	416	4	043586
3	1877.5	87.2	416	4	095657
4	1748	81.2	4	043585	043585 homo sapien
5	343	15.9	389	5	08MMI2
6	317.5	14.8	533	5	08MYI0
7	305.5	14.2	177	4	096HR6
8	288.5	13.4	494	5	09VDI1
9	283	13.2	889	4	08IM22
10	283	13.2	906	4	014526
11	272	12.6	376	4	096CF5
12	257	11.9	785	11	08C012
13	256	11.9	873	11	08K285
14	237.5	11.0	476	6	08HXCS
15	228	10.6	506	4	08IV19
16	223	10.4	471	5	08MMW4

17	216	10.0	1133	5	08INN8	08inn8 drosophila
18	207.5	9.6	968	5	09TXS0	09txs0 caenorhabdi
19	204.5	9.5	1112	5	09VHC4	09vhc4 drosophila
20	199.5	9.3	1011	5	061639	061639 drosophila
21	199.5	9.3	1014	5	08IN02	08inn2 drosophila
22	199.5	9.3	1094	5	061618	061618 drosophila
23	199.5	9.3	1097	5	09VIF7	09vif7 drosophila
24	192	8.9	546	11	08BPR8	08bpr8 mus musculu
25	191.5	8.9	416	5	007839	007839 echinococcu
26	191.5	8.9	547	11	099L10	099l10 mus musculu
27	191	8.9	1734	11	09J109	09j109 rattus norv
28	191	8.9	1793	11	09J1R0	09j1r0 rattus norv
29	190	8.8	592	4	09H8H8	09h8h8 homo sapien
30	190	8.8	620	11	08R511	08r511 rattus norv
31	187.5	8.7	679	4	096RU3	096ru3 homo sapien
32	186.5	8.7	547	11	097531	097531 rattus norv
33	186	8.6	905	4	09BX64	09bx64 homo sapien
34	185.5	8.6	2427	5	021408	021408 caenorhabdi
35	183	8.5	378	13	08AVN0	08avn0 xenopus lae
36	183	8.5	951	5	016647	016647 caenorhabdi
37	179.5	8.3	433	11	062418	062418 mus musculu
38	179.5	8.3	439	4	096K74	096k74 homo sapien
39	179.5	8.3	701	4	09NXX8	09nxx8 homo sapien
40	179	8.3	740	11	08K3Y2	08k3y2 mus musculu
41	178.5	8.3	471	11	070419	070419 rattus norv
42	178	8.3	432	11	08BH56	08bh56 mus musculu
43	178	8.3	509	11	0921L6	0921l6 mus musculu
44	178	8.3	509	11	08BNM5	08bnm5 mus musculu
45	177.5	8.2	426	5	007840	007840 echinococcu

ALIGNMENTS

RESULT 1
P97814 PRELIMINARY; PRT: 415 AA.
ID P97814
AC P97814;
DT 01-MAY-1997 (Tremblere, 03, Created)
DT 01-MAY-1997 (Tremblere, 03, Last sequence update)
DT 01-MAR-2003 (Tremblere, 23, Last annotation update)
DE PEST phosphatase interacting protein.
GN PSTRIP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Lasky L.;
RT "PEST Phosphatase Interacting Protein (PSTRIP1).";
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL: U87814; AAB48483.1;
DR HSSP: Q60631; IGBO.
DR MGD: MGI:1321396; Pstpip1.
DR InterPro: IPR001060; Cdc15_Fes_CIP4.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00611; FCH; 1.
DR Pfam: PF00018; SH3; 1.
DR PRINTS: PR00452; SH3DOMAIN.
DR Prodom: PD000066; SH3; 1.
DR SMART: SM00055; FCH; 1.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS50133; FCH; 1.
DR PROSITE: PS50002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 415 AA; 47590 MW; 16C0329284D2739C CRC64;

Query Match 100.0%; Score 2152; DB 11; Length 415;
Best Local Similarity 100.0%; Pred. No. 1.8e-134;
Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MMAOLFQDAFWCRDFTAHGTEVLLQRLLDGRKMKCDVEELLRQRAQAEERYGKELVQI 60
DB 1 MMAOLFQDAFWCRDFTAHGTEVLLQRLLDGRKMKCDVEELLRQRAQAEERYGKELVQI 60
QY 61 ARKAGGOTEMNSLRTSPSLKQOETENVSASHIOLALALREELRSLEEFEROKORRKYE 120
DB 61 ARKAGGOTEMNSLRTSPSLKQOETENVSASHIOLALALREELRSLEEFEROKORRKYE 120
QY 121 AIMDRVQSKSLVYKTKMESKRAYDOKCRDADDAEQAFERVSANGHOKOVEKSONKAKOC 180
DB 121 AIMDRVQSKSLVYKTKMESKRAYDOKCRDADDAEQAFERVSANGHOKOVEKSONKAKOC 180
QY 181 KESATEAEERVRQNIQELERARTEMEQEHRTTCEAFQLOEFDRLTLRNALVHNCNQLSM 240
DB 181 KESATEAEERVRQNIQELERARTEMEQEHRTTCEAFQLOEFDRLTLRNALVHNCNQLSM 240
QY 241 QCVKDELYEEVRLTLEGCDVGDINGFIQSKSTGRPPAPVPYQNYIDREVTPLIGSPS 300
DB 241 QCVKDELYEEVRLTLEGCDVGDINGFIQSKSTGRPPAPVPYQNYIDREVTPLIGSPS 300
QY 301 IOPSCGVTKRFGSLHSGPKTTPSAPASTETLTPPERNELVYASIEVQATQGNLNSA 360
DB 301 IOPSCGVTKRFGSLHSGPKTTPSAPASTETLTPPERNELVYASIEVQATQGNLNSA 360
QY 361 ODYRALYDYTAONSDDELISAGDILAVILEGEGMWTVERNGORGFVPGSYLEKL 415
DB 361 ODYRALYDYTAONSDDELISAGDILAVILEGEGMWTVERNGORGFVPGSYLEKL 415

RESULT 2
ID 043586 PRELIMINARY: PRT: 416 AA.
AC 043586;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE CD2 binding protein 1 long form (Proline-serine-threonine phosphatase
interacting protein 1).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_Taxid=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=99077800; PubMed=9857189;
RA Li J., Nishizawa K., An W., Hussey R.E., Lialios F.E., Salgia R.,
Sunder-Plassmann R., Reinherz E.L.;
RT "A Cdc15-like adaptor protein (CD2BP1) interacts with the CD2
cytoplasmic domain and regulates CD2-triggered adhesion."
RL EMBL J. 17:7320-7336(1998).
RN (2)
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL: AF038603; AAD11959.1;
DR EMBL: BC008602; AA08602.1;
DR HSSP: P07751; 1BK2.
DR Genew; HGNC:9580; PSTPIP1.
DR InterPro: IPR001060; Cdc15_Fes_CIP4.
DR Pfam: PF00611; FCH; 1.
DR PRINTS: PR00452; SH3DOMAIN.
DR PRODOM: PD000066; SH3; 1.
DR SMART: SM00055; FCH; 1.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS50133; FCH; 1.
DR PROSITE: PS50002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 416 AA: 47591 MW: 97818150B3D5D600 CRC64:

```

```

Query Match 88.9%; Score 1913.5; DB 4; Length 416;
Best Local Similarity 88.0%; Pred. No. 1.1e-118;
Matches 366; Conservative 20; Mismatches 29; Indels 1; Gaps 1.

QY 1 MMAOLFQDAFWCRDFTAHGTEVLLQRLLDGRKMKCDVEELLRQRAQAEERYGKELVQI 60
DB 1 MMAOLFQDAFWCRDFTAHGTEVLLQRLLDGRKMKCDVEELLRQRAQAEERYGKELVQI 60
QY 61 ARKAGGOTEMNSLRTSPSLKQOETENVSASHIOLALALREELRSLEEFEROKORRKYE 120
DB 61 ARKAGGOTEMNSLRTSPSLKQOETENVSASHIOLALALREELRSLEEFEROKORRKYE 120
QY 121 AIMDRVQSKSLVYKTKMESKRAYDOKCRDADDAEQAFERVSANGHOKOVEKSONKAKOC 180
DB 121 AIMDRVQSKSLVYKTKMESKRAYDOKCRDADDAEQAFERVSANGHOKOVEKSONKAKOC 180
QY 181 KESATEAEERVRQNIQELERARTEMEQEHRTTCEAFQLOEFDRLTLRNALVHNCNQLSM 240
DB 181 KESATEAEERVRQNIQELERARTEMEQEHRTTCEAFQLOEFDRLTLRNALVHNCNQLSM 240
QY 241 QCVKDELYEEVRLTLEGCDVGDINGFIQSKSTGRPPAPVPYQNYIDREVTPLIGSPS 300
DB 241 QCVKDELYEEVRLTLEGCDVGDINGFIQSKSTGRPPAPVPYQNYIDREVTPLIGSPS 300
QY 301 IOPSCGVTKRFGSLHSGPKTTPSAPASTETLTPPERNELVYASIEVQATQGNLNSA 360
DB 301 IOPSCGVTKRFGSLHSGPKTTPSAPASTETLTPPERNELVYASIEVQATQGNLNSA 360
QY 361 ODYRALYDYTAONSDDELISAGDILAVILEGEGMWTVERNGORGFVPGSYLEKL 415
DB 361 ODYRALYDYTAONSDDELISAGDILAVILEGEGMWTVERNGORGFVPGSYLEKL 415

RESULT 3
ID 095657 PRELIMINARY: PRT: 416 AA.
AC 095657;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE PEST phosphatase Interacting protein homolog.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_Taxid=9606;
RN (1)
RP SEQUENCE FROM N.A.
RA Wilson L.A., Fields D., Cruz L., Lasky L., Friesen J.,
Siminovitch K.A.;
RT "The human homologue of mouse PTP-PIP Interactor protein."
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL: U94778; AAD00762.1;
DR HSSP: Q0631; 1GBO.
DR InterPro: IPR001060; Cdc15_Fes_CIP4.
DR Pfam: PF00611; FCH; 1.
DR PRINTS: PR00452; SH3DOMAIN.
DR PRODOM: PD000066; SH3; 1.
DR SMART: SM00055; FCH; 1.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS50133; FCH; 1.
DR PROSITE: PS50002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 416 AA: 47610 MW: 7CED24A2E1A7EEFF CRC64:

Query Match 87.2%; Score 1877.5; DB 4; Length 416;
Best Local Similarity 86.1%; Pred. No. 2.6e-116;
Matches 358; Conservative 23; Mismatches 34; Indels 1; Gaps 1.

QY 1 MMAOLFQDAFWCRDFTAHGTEVLLQRLLDGRKMKCDVEELLRQRAQAEERYGKELVQI 60

```

```

Db      1 MMOLQFKDAFWCRDFAHTGYEVLLQRLDGRKMKCDMEELLQRAHAERYKGLVQI 60
QY      61 ARAGGOTEMNSLRTPSDLSKQOTENYNSAHQIALALREELRSLFEFRERQKORRKYE 120
Db      61 ARAGGOTEMNSLRTPSDLSKQOTENYNSAHQIALALREELRSLFEFRERQKORRKYE 120
QY      121 AINDRYOKSKLSLYKKTMESEKKAIDOKCRADDADEQAFERYNSANGHOKOYKESONKAKOC 180
Db      121 AINDRYOKSKLSLYKKTMESEKKAIDOKCRADDADEQAFERYNSANGHOKOYKESONKAKOC 180
QY      121 AINDRYOKSKLSLYKKTMESEKKAIDOKCRADDADEQAFERYNSANGHOKOYKESONKAKOC 180
Db      121 AINDRYOKSKLSLYKKTMESEKKAIDOKCRADDADEQAFERYNSANGHOKOYKESONKAKOC 180
QY      181 KESATEAERYRONIEOLERARTEMEOEHRITCAFOLOEFDRLTILRNALWVHNCNLSM 240
Db      181 KESATEAERYRONIEOLERARTEMEOEHRITCAFOLOEFDRLTILRNALWVHNCNLSM 240
QY      241 QCVKDELYEVRLLTEGCGVDGDIINGFIOSKSTGRPEPAPVPYQNYNDREVPPLISPS 300
Db      241 QCVKDELYEVRLLTEGCGVDGDIINGFIOSKSTGRPEPAPVPYQNYNDREVPPLISPS 300
QY      301 IQPSCGVYIKRSGILHSGSPKTTT-SAPAASETITLTPPERNELYASIEVQATQGNLSS 359
Db      301 IQPSCGVYIKRSGILHSGSPKTTT-SAPAASETITLTPPERNELYASIEVQATQGNLSS 359
QY      360 AODYRALYDYTAQNSDELISAGDILAVILLEGEDGMMTVERNRGOFVPGSYLEKL 415
Db      360 AODYRALYDYTAQNSDELISAGDILAVILLEGEDGMMTVERNRGOFVPGSYLEKL 415
QY      361 AQEYRAFYDYTAQNPDELISAGDILEVILEGEDGMMTVERNRGOFVPGSYLEKL 416
Db      361 AQEYRAFYDYTAQNPDELISAGDILEVILEGEDGMMTVERNRGOFVPGSYLEKL 416

RESULT 4
ID      043585 PRELIMINARY; PRT; 397 AA.
AC      043585;
DT      01-JUN-1998 (TREMBLrel. 06, Created)
DT      01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT      01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE      CD2 binding protein 1 short form.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN      NCBI_TaxID=9606;
RP      [1]
RA      LEHMANN R., BAUNGART C., FARRA G., APRIL J.F., GUIGO R., KUMPF K.,
RT      "Sequence and Analysis of Chromosome 2 of Dictyostellium.";
RL      Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RC      -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR      EMBL: AC123513; AAM44380.1; -
DR      HSSP: P0751; 1BK2.
DR      InterPro: IPR001060; Cdc15_Fes_CIP4.
DR      InterPro: IPR001452; SH3.
DR      Pfam: PF00611; FCH; 1.
DR      PRINTS: PR00452; SH3DOMAIN.
DR      ProDom: PD000066; SH3; 1.
DR      SMART: SM00055; FCH; 1.
DR      SMART: SM00326; SH3; 1.
DR      PROSITE: PS50133; FCH; 1.
DR      PROSITE: PS50002; SH3; 1.
KW      SH3 domain.
SQ      SEQUENCE 397 AA; 45353 MW; 54D64A3AAE16A2PC CRC64;

Query Match      81.28; Score 1748; DB 4; Length 397;
Best Local Similarity 82.0%; Pred. No. 8.9e-108;
Matches 341; Conservative 20; Mismatches 35; Indels 20; Gaps 2;

QY      1 MAAOLFRAFWCRDFAHTGYEVLLQRLDGRKMKCDVEELLQRAHAERYKGLVQI 60
Db      1 MAAOLFRAFWCRDFAHTGYEVLLQRLDGRKMKCDVEELLQRAHAERYKGLVQI 60
QY      61 ARAGGOTEMNSLRTPSDLSKQOTENYNSAHQIALALREELRSLFEFRERQKORRKYE 120
Db      61 ARAGGOTEMNSLRTPSDLSKQOTENYNSAHQIALALREELRSLFEFRERQKORRKYE 120

```

```

Db      61 ARAGGOTEMNSLRTPSDLSKQOTENYNSAHQIALALREELRSLFEFRERQKORRKYE 120
QY      121 AINDRYOKSKLSLYKKTMESEKKAIDOKCRADDADEQAFERYNSANGHOKOYKESONKAKOC 180
Db      121 AINDRYOKSKLSLYKKTMESEKKAIDOKCRADDADEQAFERYNSANGHOKOYKESONKAKOC 180
QY      121 AINDRYOKSKLSLYKKTMESEKKAIDOKCRADDADEQAFERYNSANGHOKOYKESONKAKOC 180
Db      121 AINDRYOKSKLSLYKKTMESEKKAIDOKCRADDADEQAFERYNSANGHOKOYKESONKAKOC 180
QY      181 KESATEAERYRONIEOLERARTEMEOEHRITCAFOLOEFDRLTILRNALWVHNCNLSM 240
Db      181 KESATEAERYRONIEOLERARTEMEOEHRITCAFOLOEFDRLTILRNALWVHNCNLSM 240
QY      241 QCVKDELYEVRLLTEGCGVDGDIINGFIOSKSTGRPEPAPVPYQNYNDREVPPLISPS 300
Db      241 QCVKDELYEVRLLTEGCGVDGDIINGFIOSKSTGRPEPAPVPYQNYNDREVPPLISPS 300
QY      301 IQPSCGVYIKRSGILHSGSPKTTT-SAPAASETITLTPPERNELYASIEVQATQGNLSS 359
Db      301 IQPSCGVYIKRSGILHSGSPKTTT-SAPAASETITLTPPERNELYASIEVQATQGNLSS 359
QY      360 AODYRALYDYTAQNSDELISAGDILAVILLEGEDGMMTVERNRGOFVPGSYLEKL 415
Db      360 AODYRALYDYTAQNSDELISAGDILAVILLEGEDGMMTVERNRGOFVPGSYLEKL 415
QY      361 AQEYRAFYDYTAQNPDELISAGDILEVILEGEDGMMTVERNRGOFVPGSYLEKL 416
Db      361 AQEYRAFYDYTAQNPDELISAGDILEVILEGEDGMMTVERNRGOFVPGSYLEKL 416

RESULT 5
ID      08MML2 PRELIMINARY; PRT; 389 AA.
AC      08MML2;
DT      01-OCT-2002 (TREMBLrel. 22, Created)
DT      01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT      01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE      Synaptic dynamin-associated protein 118B. 6/101.
OS      Dictyostelium discoideum (Slime mold).
OC      Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
OX      NCBI_TaxID=44689;
RN      [1]
RA      LEHMANN R., BAUNGART C., FARRA G., APRIL J.F., GUIGO R., KUMPF K.,
RT      "Sequence and Analysis of Chromosome 2 of Dictyostellium.";
RL      Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RC      -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR      EMBL: AC123513; AAM44380.1; -
DR      HSSP: P0751; 1BK2.
DR      InterPro: IPR001060; Cdc15_Fes_CIP4.
DR      InterPro: IPR001060; Neu_cyt_Fact_2.
DR      InterPro: IPR001452; SH3.
DR      Pfam: PF00611; FCH; 1.
DR      PRINTS: PR00452; SH3DOMAIN.
DR      ProDom: PD000066; SH3; 1.
DR      SMART: SM00055; FCH; 1.
DR      SMART: SM00326; SH3; 1.
DR      PROSITE: PS50133; FCH; 1.
DR      PROSITE: PS50002; SH3; 1.
KW      SH3 domain.
SQ      SEQUENCE 389 AA; 44258 MW; 64D5594DAF14909B CRC64;

Query Match      15.9%; Score 343; DB 5; Length 389;
Best Local Similarity 23.7%; Pred. No. 6.7e-15;
Matches 99; Conservative 85; Mismatches 194; Indels 40; Gaps 8;

QY      6 QFRDAFWCRDFAHTGYEVLLQRLDGRKMKCDVEELLQRAHAERYKGLVQI 65
Db      6 QFRDAFWCRDFAHTGYEVLLQRLDGRKMKCDVEELLQRAHAERYKGLVQI 65
QY      4 QFKDNFW-----GNCFETTEKRMNOCSTESTRLFLIKERAALEENYKSLQKLKSTS 58
Db      4 QFKDNFW-----GNCFETTEKRMNOCSTESTRLFLIKERAALEENYKSLQKLKSTS 58
QY      66 GQTEMNSLRTPSDLSKQOTENYNSAHQIALALREELRSLFEFRERQKORRKYE-VYAI 122
Db      66 GQTEMNSLRTPSDLSKQOTENYNSAHQIALALREELRSLFEFRERQKORRKYE-VYAI 122
QY      59 QLIBYGLTRDAWVGVRKASIVVHLELQKLEKTDIVAPPSKFSQKVKKNFLYDAY 118
Db      59 QLIBYGLTRDAWVGVRKASIVVHLELQKLEKTDIVAPPSKFSQKVKKNFLYDAY 118
QY      123 MDRVOKSKLSLYKKTMS-----KRAYDOKCRADDADEQAFERYNSANGHOKOYKESONKA 177
Db      123 MDRVOKSKLSLYKKTMS-----KRAYDOKCRADDADEQAFERYNSANGHOKOYKESONKA 177

```

```

Db      11  -----KLNKREKDMESSITTKRAVDYDSKQAEITAITMETAKNTTAAEVGKIQSL 17
QY      178  KCKESATEAERYRNINQILERARTEMQEHRTTCEAFQLOEPRLLITLNALNVHNO 23
Db      172  OKTORASSSEDDYDRSVNKLMSYQPTWEDKYSNHTLQTLPEERIDYIKVQLEKYGA 23
QY      238  LSMQCVKDELYEAVRLTEGCDVEEDINGFQSKSTGREPAPPYONYVDREVTPLIG 29
Db      232  IKSPVDTLEPNNLNVNITQIDKLEDIHCFRSERTGEXKPPQGLSPFGKSSDDIQ 29
QY      298  SPSIDPCGYIKRFSGLHGSPKPTTPSAPAASTELLTPPERNELVYASIEVQATQCLN 35
Db      292  NKA-----SYSAPLTSSVSSNSLTSSYSNASTATTPP-----APRSPIPLMS 33
QY      358  SSADQYRALYDTAONSDLDISAGDILAVILIEGDDGMWTVERNORGFEVGSYLEK 41
Db      333  KKMQ-AKALYDYGSA TELDFPAGDITITLDEDSGMRGELGRIDILYSNICEPT 36

```

RESULT 6

ID	PRELIMINARY:	PRT:	533 AA.
AC	OBMY10:		
DT	01-OCT-2002 (TREMBLrel. 22, Created)		
DT	01-OCT-2002 (TREMBLrel. 22, Last sequence update)		
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)		
DT	Sequence 3 from Patent WO0075321. 6/101.		
OS	Dictyostelium discoideum (Slime mold).		
OC	Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.		
OX	NCBI_taxid=44689;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	SPRAIN-AX4:		
RA	Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P., Lehmann R., Baumgart C., Parra G., Apil J.F., Gujro R., Kumpf K., Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.: "A Sequence and Analysis of Chromosome 2 of Dictyostelium". Submitted (MAY-2002) to the EMBL/Genbank/DBJ databases."		
RL	1- SIMILARITY: CONTAINS 2 SH3 DOMAINS.		
DR	EMBL: AC111072; AAM33153.1; -		
DR	InterPro: IPR001060; Cdci15_Fes_CIP4.		
DR	InterPro: IPR001452; SH3.		
DR	Pfam: PF00611; FCH; 1.		
DR	PRINTS: PR00452; SH3DOMAIN.		
DR	ProDom: PD000066; SH3; 1.		
DR	SMART: SM00055; FCH; 1.		
DR	SMART: SM00326; SH3; 2.		
DR	PROSITE: PS01033; FCH; 1.		
DR	PROSITE: PS50002; SH3; 2.		
KW	SH3 domain.		
SO	SEQUENCE 533 AA; 60356 MW; 3EE7E888B6317FB0 CRC64;		

Query March	14.8%	Score 317.5;	DB 5,	Length 533;
Best Local Similarity	21.4%	Pred. No. 4.8e-13;		
Matches 105; Conservative	97;	Mismatches 182;	Indels 107;	Gaps 15

[illegible]

Db 226 ----- : : : : :
MEVITONELVPAAVITIHNNIKKIGISIDREDLONLYILVTMSGAKPPPEAYE 279

QY 286 NY-----YDREVTPLIGSPSTPOSCGVYK-----RFSGLLGSP-- 319

Db 280 PLOGGGGFAIVSSSSNSLNITSRKSGELNGGSIONGASITSSPQOYQYVIMDHTPPQ 332

QY 320 -----KTPPSAPAASTETLTPTPERNEL-----VY 344

Db 340 NIITQHOQSNNNNTNNNSNYMTTPPP-----QFPPQOOLPQPTOLNPPPPIS 392

QY 345 ASIEVQATQGNLNSAODYRALYDYTAQNSDELDISADILAVILEGEGMW---ITYERN 401

Db 393 LSRKNDSSNSINSNGEIVRALYDYNAATEENIEISFRANALIKVYLREDSGWMQOMVIGES 452

QY 402 GORGFPVGSYL 412

Db 453 DRIGVFPFSNFI 463

RESULT 7

ID	096HR6;	PRELIMINARY;	PRT;	177 AA.
AC	096HR6;			
DT	01-DEC-2001 (TrEMBLrel. 19, Created)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)			
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)			
DE	Similar to proline-serine-threonine phosphatase-interacting protein 2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Eye;			
RA	Strausberg R.;			
RL	Submitted (May-2001) to the EMBL/Genbank/DBJ databases.			
DR	EMBL; BC008216; AA008216.1; -			
DR	Interpro; IPR001060; Cdc15_Fes_CIP4.			
DR	Pfam; PF00611; FCH; 1.			
DR	SMART; SM00055; FCH; 1.			
DR	PROSITE; PS01033; FCH; 1.			
SQ	SEQUENCE 177 AA; 20422 MW; EC6587EA488F56 CRC64;			

Query March	14.2%	Score 305.5	DB 4	Length 177
Best Local Similarity	48.8%	Pred. No. 7.7e-13		
Matches 59; Conservative	26	Mismatches 35	Indels 1	Gaps 1

```

QY      2 M A O L G R D A F M W C D E P T A H G T E V L L Q R L D G K M C D V E E L L R R A Q A E E R Y G K E I V O I A 61
Dd      1 M T R S L F K G M F W S A D I I S T I G Y D N I T Q I H L N G K R K C K E F D P L K R A A I E E R Y G D L N L S 60

QY      62 R K A G - G O T E M N S L I R T S F D S L K O O T E N V G S A H I T O L A L R E E L R S L E E F R E R O K E O R K Y E 120
Dd      61 R K R P G Q S E I N I L K R A L E F Y K Q V D N V A Q C H I Q L A Q S L R E A R K M E E R E N O K I Q R R K M E 120

QY      121 A 121
Dd      121 S 121

```

RESULT 8

ID	OSVDIL	PRELIMINARY;	PRT;	494 AA.
AC	OSVDIL; OS8Z24;			
DT	01-MAR-2000 (TREMBLrel. 13, Created)			
DT	01-OCT-2002 (TREMBLrel. 22, Last sequence update)			
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)			
DE	CG15594 protein (TMD46528p).			
GN	STNDAPIN OR CG15594.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota			

RESULT 9
O81W22 PRELIMINARY: PRT: 889 AA.
AC O81W22
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DE Similar to RIKEN CDNA 3322402E17 gene.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Blood;
RA Strausberg R.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC041130; AAAH41130.1;
SQ SEQUENCE 889 AA; 96860 MW; 8CD41F24F42256D2 CRC64;

Query Match 13.2%; Score 283; DB 4; Length 889;
Best Local Similarity 24.3%; Pred. No. 1.7e-10;
Matches 93; Conservative 90; Mismatches 148; Indels 52; Gaps 15;

OY 7 FRDAFWCRDFTAHNGEYVLLQRLDGRKMKDVEELLROAQAEEYRGKELVQIARKAG 66
DB 4 FGEHFWGE---KNHGFVLYHSVKQGPSTKELADFIETRYETYSKMAKLSKLAN 60
OY 67 QTEMNSLRTSFSDSLKQOTENVGSAHQLALRELSLEEFREKQKRYE---AI 122
DB 61 GTPMGTRPAPLMEYFRVSSDKALCHLELTRKLDLIKVLRYGEGEOLKTHKCKEEVST 120
OY 123 MDVQKSLSYKKT--ESKRAYDQKRDADDAQAFAERYASANGHOKOVEKSONKAKCK 181
DB 121 LDAOV--VLSGVSQLPKSRRENYLNRCMD---QERLRREST--QKEMDAETKTKAA 172
OY 182 ESATEAEVRYRONIEOLARTEWEOHRTTCEAFQLOEFDRLTLRNALVWNCNLSMQ 241
DB 173 ES-----LRSVEKYNSARADFEQKMLDSALRFQAMEETHLRHMKALLGSYHVEDI 225
OY 242 CVKDELIEEYRLTECGDVEGDIINGFLOSSTGREGPPAPVYQNYDREVTPLIGSPST 301
DB 226 HVOIGVHEEFKONIEVSEMLLRKFAESKGTGREGPPGLDFEAY-----SAAL 276
OY 302 QPSCGVIKRPSGL---LHG--SPKTPPSAPASTETLTP---TPERNELVASTIEVQAT 352
DB 277 QEA---MKRLRGAKAFRLPGLSREREPPEPAVDFLEPDSGTCPEVDEGF--TVRPDVT 332
OY 353 QGNINSSAODYRALYDTAQNDS 375
DB 333 Q---NSTAEPSSR---FSSSDSD 348

RESULT 10
O14526 PRELIMINARY: PRT: 906 AA.
AC O14526
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Hypothetical protein KIAA0290 (Fragment).
GN KIAA0290.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Brain;
RA Ohara O., Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Nomura N.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL: AB006628; BAA22959.1; -
DR InterPro: IPR001060; Cdc15_Fes_C1P4.
DR InterPro: IPR001917; NHRtransf_2.
DR Pfam: PF00611; FCH; 1.
DR SMART: SM00055; FCH; 1.
DR PROSITE: PS00599; AA_TRANSFER_CLASS_2; 1.
DR PROSITE: PS01333; FCH; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 906 AA; 98738 MW; DABD36C2DA02F567 CRC64;

Query Match 13.2%; Score 283; DB 4; Length 906;
Best Local Similarity 24.3%; Pred. No. 1.7e-10;
Matches 93; Conservative 90; Mismatches 148; Indels 52; Gaps 15;

OY 7 FRDAFWCRDFTAHNGEYVLLQRLDGRKMKDVEELLROAQAEEYRGKELVQIARKAG 66
DB 21 FGEHFWGE---KNHGFVLYHSVKQGPSTKELADFIETRYETYSKMAKLSKLAN 77
OY 67 QTEMNSLRTSFSDSLKQOTENVGSAHQLALRELSLEEFREKQKRYE---AI 122
DB 78 GTPMGTRPAPLMEYFRVSSDKALCHLELTRKLDLIKVLRYGEGEOLKTHKCKEEVST 137
OY 123 MDVQKSLSYKKT--ESKRAYDQKRDADDAQAFAERYASANGHOKOVEKSONKAKCK 181
DB 138 LDAOV--VLSGVSQLPKSRRENYLNRCMD---QERLRREST--QKEMDAETKTKAA 189
OY 182 ESATEAEVRYRONIEOLARTEWEOHRTTCEAFQLOEFDRLTLRNALVWNCNLSMQ 241
DB 190 ES-----LRSVEKYNSARADFEQKMLDSALRFQAMEETHLRHMKALLGSYHVEDI 242
OY 242 CVKDELIEEYRLTECGDVEGDIINGFLOSSTGREGPPAPVYQNYDREVTPLIGSPST 301
DB 243 HVOIGVHEEFKONIEVSEMLLRKFAESKGTGREGPPGLDFEAY-----SAAL 293
OY 302 QPSCGVIKRPSGL---LHG--SPKTPPSAPASTETLTP---TPERNELVASTIEVQAT 352
DB 294 QEA---MKRLRGAKAFRLPGLSREREPPEPAVDFLEPDSGTCPEVDEGF--TVRPDVT 349
OY 353 QGNINSSAODYRALYDTAQNDS 375
DB 350 Q---NSTAEPSSR---FSSSDSD 365

RESULT 11
O96CF5 PRELIMINARY: PRT: 376 AA.
AC O96CF5
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Bladder;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC014311; AAAH4311.1; -
DR InterPro: IPR001060; Cdc15_Fes_C1P4.
DR Pfam: PF00611; FCH; 1.
DR SMART: SM00055; FCH; 1.
DR PROSITE: PS01333; FCH; 1.
KW Hypothetical protein.
SQ SEQUENCE 376 AA; 43190 MW; 94279BECBEC027F CRC64;

Query Match 12.6%; Score 272; DB 4; Length 376;
Best Local Similarity 24.0%; Pred. No. 3.2e-10;
Matches 83; Conservative 76; Mismatches 131; Indels 56; Gaps 11;

QY 7 FRDAFWCRDFTAHNGEYVLLQRLDGRKCKDVEELLRQRAQAEERYKELVOIARKAGG 66
 DB 4 FENFWME---KNSGFDVLNNMHGQISTKEIADPVRERATIEAVRSRSTKLAKASN 60
 QY 67 QTEANSLSRTSFDLSKOOTENNGSAHIOLALRELSLEEFREKOEKRYE---AI 122
 DB 61 YSOLGTFAPVWDVKTSTKLANCHLDVLRKLOELIKEVOKYGEQVSKHKTKEEVAAGT 120
 QY 123 MDRVOKRLSLYKTKMESKKAYDOKCRADADAQAEFEVSANG-HOKVEKSONKAKOCK 181
 DB 121 LEAVOTIO-STLOALQSKENYNAK---VEQ--ERLKEGATQREIERAAVSKK-- 170
 QY 182 ESATEARVYRONTEOLERARTEWEOEHRTTCEAFOLQEFRLTLIRNALWVHCNLSMQ 241
 DB 171 --ATDTYKLY---VEKALAKADEQOKTEFAQKPODEEHLHIKELISLSNAIKEI 225
 QY 242 CVKDELYEVRRLTEGCDVEGDIINGPIQSKSTGREPPAPVYONYDREVTPLIGSPSI 301
 DB 226 HLOIGOVHEEFINNMANTVYESLIQKFAESKGTGERGLIEFE---ECDTASAVEGI 280
 QY 302 QPS-----CGVIR-----RFGSLHSP 319
 DB 281 KPRKRTFALPGITIKKEKDAESVSTVYVAVQMGWRDLGLHSP 326

RESULT 12

08C012 PRELIMINARY; PRT: 785 AA.
 ID 08C012
 AC 08C012;
 DT 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Hypothetical cell division control protein 15.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Thymus;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.
 RL Nature 420:563-573.(2002).
 DR EMBL, AK031041; BAC27226.1;
 KW Hypothetical protein.
 SQ SEQUENCE 785 AA; 85852 MW; 3B77B558A078E36C CRC64;

Query Match 11.9%; Score 257; DB 11; Length 785;
 Best Local Similarity 23.0%; Pred. No. 7.6e-09;
 Matches 89; Conservative 87; Mismatches 147; Indels 64; Gaps 15;

QY 32 GRKCKDVEELLRQRAQAEERYKELVOIARKAGGTEMSLRTSFDLSKOOTENNGSAH 91
 DB 4 GQISTKELADPVRERATIEAVRSRSTKLAKASN 63
 QY 92 IOLALIRELSLEEFREKOEKRYE---AIDRVOKSKLSYKTKMESKKAYDOK 147
 DB 64 LDVLRKLOELIKEVOKYGEQVSKHKTKEEVAAGTLEAVQIO-NIQAOLQSKENYNAK 122
 QY 148 CRADADAQAEFEVSANG-HOKVEKSONKAKOCKESTEAERYRONTEOLERARTEW 206
 DB 123 C-----VEQ--ERLKEGATQREIERAAVSKK---ATDYKLY---VEKALAKA 168
 QY 207 QEHRTTCEAFOLQEFRLTLIRNALWVHCNLSMQCVDELYEVRRLTEGCDVEGDI 266
 DB 169 QKMETAQOKPODIETHLHIKELISLSNAVKEHLOIGOVHEEFINNMANTVYESLIQ 228
 QY 267 GFIOKSTGREPPAPVYONYDREVTPLIGSPSIOPS-----CGVIRFGSLHSP 319
 DB 229 KFAESKGTGERGLIEFE---ECDPASAVEGIKPRKRTFALPGITIKK-----E 275

QY 320 KTFPSAPASPTLTPPERNELVYASTEVQATQGNL-----NSSAOD-----YRA 365
 DB 276 KDAESVECPDADSLN-IPVDDEGF-STKPEANQDTEKNHPIYSSDSDEDEPKRRI 333
 QY 366 -----LYDTAONSDELDISAGDI 384
 DB 334 EIKPAHPNNLHHTTASLDELKVSIGNI 360

RESULT 13

08K285 PRELIMINARY; PRT: 873 AA.
 ID 08K285
 AC 08K285;
 DT 01-OCT-2002 (TREMblrel. 22, Created)
 DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE RIKEN CDNA 332402E17 gene.
 GN 332402E17RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (JUN-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL, BC032207; AAH32207.1;
 DR MGD; MGI:1921265; 3322402E17RIK.
 DR InterPro; IPR001060; Cds15_Fes_C1P4.
 DR InterPro; IPR001917; NHRtransf_2.
 DR Pfam; PF00611; FCH; 1.
 DR SMART; SM00055; FCH; 1.
 DR PROSITE; PS00599; AA_TRANSFR_CLASS_2; 1.
 DR PROSITE; PS50133; FCH; 1.
 SQ SEQUENCE 873 AA; 95151 MW; E482119EF9A57020 CRC64;

Query Match 11.9%; Score 256; DB 11; Length 873;
 Best Local Similarity 21.4%; Pred. No. 1e-08;
 Matches 92; Conservative 96; Mismatches 177; Indels 64; Gaps 13;

QY 7 FRDAFWCRDFTAHNGEYVLLQRLDGRKCKDVEELLRQRAQAEERYKELVOIARKAGG 66
 DB 4 FGEHFW---GKNHGEFVLYHGVKGPATKELAFINERANIEETYSKAMKLSASN 60
 QY 67 QTEANSLSRTSFDLSKOOTENNGSAHIOLALRELSLEEFREKOEKRYEPAIDR 125
 DB 61 GTPMGTFAPLWVEFVSSDKLALCHLELRKLIHDLKDLVLRGEBOLTKHKCKEEVLGT 120
 QY 126 VOKSKL--SLYKTKMESKKAYDOKCRADADAQAEERY--SANGHOKVEKSONKAKOCKE 182
 DB 121 VDANOVLGSGQLLPKSKRENTLSRCMD-----LERLRRENTSOKEMDKATSKKAA 173
 QY 183 SATEARVYRONTEOLERARTEWEOEHRTTCEAFOLQEFRLTLIRNALWVHCNLSMQC 242
 DB 174 S-----LRSVQDYNSARADFEIKMDSALRFQAMEAHQHKALLGSAHSEVEDTH 226
 QY 243 VKDELYEVRRLTEGCDVEGDIINGPIQSKSTGREPPAPVYONYDREVTPLIGSPSI 302
 DB 227 VOIGOVHEEFQWENVTVDMLLRKFAESKGTGERKPGPLDPDAY-----SSAALQ 277
 QY 303 PSCGVIRFGSL-----LHGSPKTPPSAPASPTLTPPERNELVYASTEVQATQGN-- 355
 DB 278 EA--MKRLRGAKAFRLPGLSRPRASVDFLESQGVPEYDDEGCFVPRDISONNGAE 334
 QY 356 ---LNSAODY-----RALDYTAONSDELDISAGDIILAVILLEGEDGW 395
 DB 335 PPRSSSDPDDEEPRKFTYHIKPAITRAVACSSAAAOURLATAGSL--ILPGPGG- 391
 QY 396 WTVERNGOR 404
 DB 392 -TMKRHSR 399

RESULT 14

Q8HXC5 PRELIMINARY; PRT; 476 AA.
 ID 08HXC5
 AC 08HXC5
 DT 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Hypothetical protein.
 OS Macaca fascicularis (Crab-eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Macaca.
 NCBI_TaxID=9541;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Frontal cortex;
 RA Hashimoto K., Osada N., Hida M., Kusuda J., Sugano S.;
 RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Frontal cortex;
 RX MEDLINE=21458551; PubMed=11574149;
 RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirata M., Suto Y.,
 RA Hirai M., Terao K., Suzuki Y., Sugano S., Hashimoto K.;
 RT "Assignment of 118 novel cDNAs of cynomolgus monkey brain to human
 RT chromosomes.";
 RL Gene 275:31-37(2001).
 DR EMBL; AB093665; BAC21639.1;
 KW Hypothetical protein.
 SO SEQUENCE 476 AA; 54374 MW; CA47BF56B3F4E86F CRC64;

Query Match

Best Local Similarity 11.0%; Score 237.5; DB 6; Length 476;
 Matches 69; Conservative 62; Mismatches 130; Indels 19; Gaps 6;

QY 9 DAFWC-RDFTAH---TGVEVLRLDGRKCKKVEELLRRAQAEEERYKELVOLARK 63
 DB 202 DTFADKDPGNGTAVAGELLQKOLGKQKQKSEFIRIRITIEDYAKNLKLSQN 261
 QY 64 AGGOTENSLRTSPSLKQOTENVSANILQALALREEL-RSELEFEROKRQKKYAI 122
 DB 262 SLASQEGSLIGAWAQVKKSLADEAVHLKPSAKHSEVEKPLMFRENFKKDKKCDH 321
 QY 123 MDVQKSLSLYKTKMESKKAYDQCKRADDAEQAFERVSANGHOKQVEKSNARAKQCKE 182
 DB 322 IADLKQOLASRYASVEKARKALTERQDLEMKTQOLETKLSNKEEDDKARRKSTQAGD 381
 QY 183 SATEARVYRONIEQLERARTMEQEHRTTCEAFQLOEPDRLTILRNALWVHCNQLSMQC 242
 DB 382 DLMRCVDLYNQ-----AQSKWPEEMVTTTLELEERLEVEREYEMIRHNL---COTQLRH 431
 QY 243 VKD---DELYEEVRLTLEGCEVEGDINGFIQSKSTGREPP 279
 DB 432 ETDMEFNOSTVEPVDQLRKVPAPKDRLELWREHKTGSIRP 471

RESULT 15

Q8IVT9 PRELIMINARY; PRT; 506 AA.
 ID 08IVT9
 AC 08IVT9
 DT 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE NOSTRIN protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Zimmermann K., Oplitz N., Dedio J., Renne C., Muller-Esterl W.,

RA Oess S.;
 RT "NOSTRIN: A protein modulating nitric oxide release and subcellular
 RT distribution of endothelial nitric oxide synthase";
 RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(2002).
 DR EMBL; AJ532842; CAD58724.1;
 SO SEQUENCE 506 AA; 57731 MW; DC993BC0D9C2AED0 CRC64;

Query Match

Best Local Similarity 10.6%; Score 228; DB 4; Length 506;
 Matches 96; Conservative 88; Mismatches 180; Indels 148; Gaps 13;

QY 22 YEVLRLDGRKCKKVEELLRRAQAEEERYKELVOLAR---RRAGOTENSLRTSPD 78
 DB 14 YKMLEFSQNGENFCKQVTSVLOQRANLEISTAKGLQKSLASLALNTRKSCVSSAWA 73
 QY 79 SLKQOTENVSANILQALALREELRSELEFEROKRQKKYAIMDRYKQ----- 129
 DB 74 WASEGKSYADLHQKLGKAI--ELEAIKPTQYVLVWQEKRRSLDNEVEKTAANLVISWN 131
 QY 130 -KLSLYKTKMESKKAYDQCKRADDAEQAFERVSANGHOKQVEKSNK-AKQKESATPA 187
 DB 132 QQIKAKKKLWSTKHEALFQLVESKQSMT-----EKEKRKLNLKLTSTKLEKED 184
 QY 188 ERYVYRONIEQLERARTMEQEHRTTCEAFQLOEPDRLTILRNALWVHCNQLSMQCKDE 247
 DB 185 ENYVQKNMAGYS-TRKMENTLENQCYQSLLEKERRIQLDCLNQLNQSQHSIFGQITTT 243
 QY 248 LYEEVRLTLEGCEVEGDINGFIQSKS-TGREPPAPVPYONY-----DREYTPILGS 298
 DB 244 CHTQIHCAISKIDIEKIDQAVMEETALSTENKSPFLTDYFEEEDPNSAMDKERRKSLK 303
 QY 299 PSI----- 301
 DB 304 PKILRLQDIEKASKQKGLERMLKTYSTSSFSQDANSQKDTAALMDENNLKLDLEANS 363
 QY 302 -----QPS--CG-----VIRFSGLLHSPKPTTP 323
 DB 364 YKLSSMLAELBQRPQSPHCNSLFRMRKHTHSYKISPFMLKRLNENVSAASSGCG 423
 QY 324 SAPASTETLTPTPEERNELYASIEVQATGNNLSSAQDYRALDYTAQNSDELIDISAGD 383
 DB 424 SNPSS---TPAP-----GAOLSSRLKALYSFOARODDELINLEKGD 463
 QY 384 ILAVILEGDEGMVTVERNQGFVPGSYLEKL 415
 DB 464 IVIIHEKKEBGMWEGSLNGKKGHPAPAYVEL 495

Search completed: July 28, 2003, 09:17:28
 Job time : 42 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: August 3, 2003, 11:41:51 ; Search time 3433 Seconds
(without alignments)
3518.108 Million cell updates/sec

Title: US-09-068-377c-1
Perfect score: 2152
Sequence: 1 MAAQLQFPADEWCRDFAHT.....WTVRRNGRGVPGVSYLEKL 415

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame.p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US09068377/runat.28072003.091316.15316/app.query.fasta.1.583
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human40.cdi -LIST=45
-DOCCALIGN=200 -THR.SCORE=DCT -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=plc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09068377.GC.GN.1.1.2456 -runat.28072003.091316.15316 -NCPU=6 -ICPU=3
-NO_MAP -LARGEQUERY -NEG.SCORES=0 -WAIT -DSPBLOCK=100 -LONLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: GenEmbl:*
2: gb_hgt:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vl:*
30: em_hgt_hum:*
31: em_hgt_fov:*
32: em_hgt_other:*
33: em_hgt_mus:*
34: em_hgt_pln:*
35: em_hgt_rod:*
36: em_hgt_mam:*
37: em_hgt_vrl:*
38: em_sy:*
39: em_hgt_hum:*
40: em_hgt_mus:*
41: em_hgt_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2152	100.0	1692	10 MM087814	U87814 Mus musculus
2	2152	100.0	2100	6 AF108626	AF108626 Sequence
3	1913.5	88.9	1428	9 AF038603	AF038603 Homo sapi
4	1913.5	88.9	1781	9 BC008602	BC008602 Homo sapi
5	1877.5	87.2	1656	9 HS094778	HS094778 Human PEST
6	1748	81.2	1382	9 AF038602	AF038602 Homo sapi
7	737.5	34.3	1498	10 BC002123	BC002123 Mus muscu
8	736.5	34.2	1005	10 MM018101	MM018101 Mus muscu
9	736	34.2	1580	6 AX058229	AX058229 Sequence
10	510	23.7	161469	2 AC116689	AC116689 Mus muscu
11	503.5	23.4	1562	9 IROEST015	AI101110 Homo sapi
12	465	21.6	173635	2 AC123432	AC123432 Rattus no
13	452.5	21.0	2591	9 AK023100	AK023100 Homo sapi
14	317.5	14.8	115489	2 AC117072	AC117072 Dictyoste
15	305.5	14.2	1461	9 BC008216	BC008216 Homo sapi
16	302.5	14.1	22500	8 SPBC11C11	AL031528 S.pombe c
17	301	14.0	3255	6 AX193116	AX193116 Sequence
18	301	14.0	3255	9 AF128536	AF128536 Homo sapi
19	299	13.9	2180	9 BC008037	BC008037 Homo sapi
20	297	13.8	3993	8 SPDCD15	X86179 S.pombe cdc
21	297	13.8	30200	8 SPAC2068	295334 S.pombe chr
22	288.5	13.4	2017	3 AY069690	AY069690 Drosophi
23	285.5	13.3	2681	9 AK056770	AK056770 Homo sapi
24	285.5	13.3	3176	9 HSM601813	AL136845 Homo sapi
25	284.5	13.2	2079	10 AF139492	AF139492 Rattus no
26	283.5	13.2	1695	9 AF242529	AF242529 Homo sapi
27	283.5	13.2	2826	10 BC023502	BC023502 Mus muscu
28	283.5	13.2	2983	6 AX193117	AX193117 Sequence
29	283.5	13.2	3217	10 AF128535	AF128535 Mus muscu
30	283.5	13.2	4214	9 AB037800	AB037800 Homo sapi
31	283.5	13.2	4294	9 HSM605614	AL834211 Homo sapi
32	283.5	13.2	4319	6 AX405634	AX405634 Sequence
33	283	13.2	2979	9 AB006628	AB006628 Homo sapi
34	283	13.2	3460	6 AX405906	AX405906 Sequence
35	281	13.1	200703	2 AC123482	AC123482 Rattus no
36	279.5	13.0	1777	10 AF242531	AF242531 Mus muscu
37	279.5	13.0	1814	10 BC003884	BC003884 Mus muscu
38	279.5	13.0	2073	10 AF139494	AF139494 Rattus no
39	279.5	13.0	4909	9 HSM603278	AL831971 Homo sapi
40	277.5	12.9	1767	10 AF149824	AF149824 Mus muscu
41	275	12.8	1311	9 BC014311	BC014311 Homo sapi
42	274	12.7	1956	10 AF139493	AF139493 Rattus no
43	273.5	12.7	1597	10 AF104402	AF104402 Rattus no
44	269	12.5	1950	10 AF139495	AF139495 Rattus no
45	268	12.5	1825	9 AF149825	AF149825 Homo sapi

RESULT 1

ALIGNMENTS

MMU087814 1692 bp mRNA linear ROD 30-OCT-2001
LOCUS MMU087814
DEFINITION Mus musculus PEST phosphatase interacting protein mRNA, complete
ACCESSION U087814
VERSION U087814.1 GI:1857711
KEYWORDS
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 1692)
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Spencer, S., Dowbenko, D., Cheng, J., Li, W., Brush, J., Utzig, S.,
Simanis, V., and Lasky, L.A.
STPPIP: a tyrosine phosphorylated cleavage furrow-associated
protein that is a substrate for a PEST tyrosine phosphatase
J. Cell Biol. 138 (4), 845-860 (1997) **Aug 25**
TITLE
JOURNAL JOURNAL
MEDLINE 97412181
PUBMED 9265651
REFERENCE 2 (bases 1 to 1692)
Lasky, L.
AUTHORS Direct Submission
JOURNAL Submitted (29-JAN-1997) Molecular Oncology, Genentech, 460 Pt. San
Bruno Blvd., San Francisco, CA 94080, USA
FEATURES
Source Location/Qualifiers
1..1692
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_type="hematopoietic progenitor cells"
274..1521
/note="Similar to S. pombe Cdc15p, a protein associated
with formation of the cleavage furrow during cytokinesis;
actin interacting protein with C-terminal SH3 domain;
STPPIP"
/codon_start=1
/product="PEST phosphatase interacting protein"
/protein_id="AAB4843.1"
/db_xref="GI:1857712"
/translation="MMAQLQFQDAFMCRRDTAHTGYEVLLRLDGRKMKDVEELR
QRAQAEERYKEIVQARKAGQTEMNLSPTSLKQENTGSAHIQIALRELL
RSLLEEFERQEKQRYKEAIVMDRQKSLSLYKTMESKAYQKCRDADQAEER
VSANGHOKQVEKSONAKQCKESATEERYRONITQLEFARTEMDEHRTTEAQL
QEDRLTILNALMVHCNOLSMOCVDEDEYEVRLTLECDVEDGIDNGTQSTGR
EPAPAPYQNYVREVTPLIGSIPSCGVIKRFSGLHSGKTPSPASASTETLT
PTPERNELVASTEQATQGNLSADYRALDYTAONSDELDISAGDILAVILLEGE
DGMWTVERNQGRFVPGSYLEKL"
BASE COUNT 409 a 443 c 526 g 314 t
ORIGIN
Alignment Scores:
Pred. No.: 8.97e-151 Length: 1692
Score: 2152.00 Matches: 415
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
US-09-068-377c-1 (1-415) x MMU087814 (1-1692)
QY 1 MetMetAlaGluLeuGlnPheArgAspAlaPheTrpCysArgAspPheThrAlaHisThr 20
DB 274 ATGATGGCCAGCTGAGTTCGAGATGCTTCTGTGCGAGGACTTCACGGCCACACA 333
QY 21 GYTTCGValLeuLeuGlnArgLeuLeuAspGlyArgLysMetCysLysAspValGlu 40
DB 334 GGGTATGAGGTGCTACTGACAGAGGCTGCTGACGCGAGAGATGCAAGGATGGAG 393
QY 41 GluLeuLeuArgGlnArgAlaGlnAlaGlnGluGluArgTyrGlyLysGluLeuValGlnIle 60
DB 394 GAGCTGCTACAGACAGAGGCCACAGCGGAGGAGAGGTACGGAGAGGCTGTGAGAT 453
QY 61 AlaArgLysAlaGlyGlyGlnThrGluMetAsnSerLeuArgThrSerPheAspSerLeu 80

DB 454 GCACCGAAGCTGTGTGCCACAGACAGATGAATTCCTGAGGACCTCTTGACTCCCTG 513
QY 81 LysGlnGlnThrGlnGluAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaArgGlu 100
DB 514 AAGCAGCAACAGAGAAATGTGGCGACATCCACACTCCAGCTGGCCCTGCGGTGAG 573
QY 101 GluLeuArgSerLeuGlnGluPheArgGluArgGlnLysGlnGlnArgLysTyrGlu 120
DB 574 GAGCTGGGAGACCTCGAGGAGTTCGAGAGACAGACAGAAAGCGACGGAAGATGAG 633
QY 121 AlaIleMetAspArgValGlnLysSerLysLeuSerLeuTyrLysTyrThrMetGluSer 140
DB 634 GCCATCATGACCTGTGTCAGAGAGCAAGTGTGCTGTACAGAGAACACATGAGATGCC 693
QY 141 LysLysAlaTyrAspGlnLysCysArgAspAlaAspAspAlaGlnGlnAlaPheGluArg 160
DB 694 AAGAGGCAATATGACAGAGAGTGCAGGATGCAAGATATCTGACAGGCTTCGAGCT 753
QY 161 ValSerAlaAsnGlyHisGlnLysGlnValGluLysSerGlnAsnLysAlaLysGlnCys 180
DB 754 GTGAGTCCCAATGGCCACAGAGCAAGTGAAGAAAGACAGCAAAAGCCAGCAGTGC 813
QY 181 LysGluSerAlaThrGlnAlaGlnArgValTyrArgGlnAsnIleGlnLeuGlnArg 200
DB 814 AAGGAGTCAACCCACAGAGCGAGAAAGATGTACAGGAAATATCGAACAACCTGGAGAGA 873
QY 201 AlaArgThrGluTrpGlnGlnGlnHisArgThrThrCysGlnAlaPheGlnLeuGlnGlu 220
DB 874 GCGAGGACCCAGTGGGAGCAGAGCAGCAGCAGCAGTACCTGTGAGCCTTCAGTGCAGGAG 933
QY 221 PheAspArgLeuThrIleLeuArgAsnAlaLeuThrValHisCysAsnGlnLeuSerMet 240
DB 934 TTTCAGCGGCTCAACCTCTCCCAATGCCCTGTGGTGCACGTGAACAGCTCTCCAG 993
QY 241 GlnCysValLysAspAspGluLeuTyrGlnGluValArgLeuThrLeuGlnGlnCysAsp 260
DB 994 CAGTGTGTCAAGATGATGAGTCTATGAGGAAGTGGGCTGACCTTAGGCGCTGTGAT 1053
QY 261 ValGlnGlyAspIleAsnGlyPheIleIleGlnSerLysSerThrGlyArgGlnProProAla 280
DB 1054 GTGGAAGGTGACATCATAGGCTTCATCAGCAGTCCAAAGCAGCAGCAGCAGCAGCAGCT 1113
QY 281 ProValProTyrGlnAsnTyrTyrAspArgGluValThrProLeuIleGlnSerProSer 300
DB 1114 CCGGTCCCTTACAGACACTACTATGACAGAGGAGTGACCCACTGATTTGGAGCCCTGAC 1173
QY 301 IleGlnProSerCysGlyValIleLysArgPheSerGlyLeuLeuHisGlySerProLys 320
DB 1174 ATCCAGCCCTCCTCGGCTGTGATAAAGAGTCTCTGGGCTGCTCATGGAAGTCCACAG 1233
QY 321 ThrThrProSerAlaProAlaAlaSerThrGlnThrLeuThrProThrProGlnArgAsn 340
DB 1234 ACCACACCTTGTGCTGCTGCTGCTTCCACAGAGACTGTGACCTCCACCCCTGAGCGGAT 1293
QY 341 GluLeuValTyrAlaSerIleGluValGlnAlaThrGlnGlyAsnLeuAsnSerSerAla 360
DB 1294 GAGTTGCTACGACATCATGCAAGTGCAGGAGCGACCCAGCAACCTTAACATCATACACC 1353
QY 361 GlnAspTyrArgAlaLeuTyrAspTyrThrAlaGlnAsnSerAspGluLeuAspIleSer 380
DB 1354 CAGGACTACCGGCGACTGTACGACACTACCTCAGCAATTTGTGATGACCTGGACATTGCC 1413
QY 381 AlaGlyAspIleLeuAlaValIleLeuGlnGluGlnAspGlyTyrPheThrValGlnArg 400
DB 1414 GCGGAGACATCCCTGCGGTATCTCGAAGGAGGAGTGGCTGGTGGATCTGTGAGCGG 1473
QY 401 AsnGlyGlnArgGlyPheValProGlySerTyrLeuGlnLysLeu 415
DB 1474 AACGCAACAGCTGGCTTGTCCCTGGGCTGCTGACTTTGGAAGAGCTC 1518
RESULT 2
AR108626
LOCUS AR108626 2100 bp DNA linear PAT 14-FEB-2001

[illegible]

Db	722	AGAGCTGTCAAGATGTATGAGCTCTACGAGGAAGTGGGGCTGAGCTGGAAAGGCTGCAGC	780
Qy	261	ValleluclYaspiIleasnGlyPheIleGlnSerIlySerThrGlyArgGluProProAla	280
Db	781	ATAGAGCCCGGACACGACGAGATTTCATCTACGAGCCCAAGAGACAGGGGCACAGACCCCGCT	840
Qy	281	ProValProThyArgIlnAsnTyrrTyAspArgGluValThrProLeuIleGlySerProSer	300
Db	841	CCGGTGGCCCTACCGAAGACTATTACGATCGGGAGGTACCCCGCTGACCAGAGGCCCTGGC	900
Qy	301	IleGlnProSerGlyGlyValIleIlyArgPheSerGlyLeuLeuHisGlySerProIlys	320
Db	901	ATPACGCGCTCGCGGCATGATAAAGGTTCTCTGACCTGCGACGAGAGCTCCAG	960
Qy	321	ThrThrPro--SerAlaProAlaAlaSerThrGluThrLeuThrProThrProGluIArg	339
Db	961	ACCACCTTGTTGGCAGCTTCGTCGCGCTCCACAGACACCCTGACCCCGCAGGAGCG	1020
Qy	340	AsnGluLeuValTyrrAlaSerIleGluValGlnAlaThrGlnIlyAsnLeuAsnSerSer	359
Db	1021	AATAGAGGTGTCCTACAGCCATCGCAGTGCAGAGATACAGGGAAACCCGCGCTCCACA	1080
Qy	360	AlaGlnAspTyrrArgAlaLeuTyrrAspTyrrThrIacIlnAsnSerAspGluLeuAspIle	379
Db	1081	GCCAGAGAGTACCGCGCGCTCTACGATTATACAGCCAGAAACCAAGATGAGCTGACCTG	1140
Qy	380	SerAlaGlyAspIleLeuAlaValIleLeuGluGlyGluAspGlyTyrPThrValGlu	399
Db	1141	TCCCGCGGAGACATCTCGAGGTGATCTCTGAAAGGAGAGATGGCTGCTGAGACTGTGAG	1200
Qy	400	ArgAsnGlyIlnArgGlyPheValProGlySerTyrrLeuGluIlyLeu	415
Db	1201	AGGAGCGGACGCGTGGCTTCGTCCCTCGATTCTTCCACTGGAGAAAGCTT	1248

RESULT	4
LOCUS	BC008602
DEFINITION	BC008602 1781 bp mRNA linear PRI 12-JUL-2001
ACCESSION	Homo sapiens, proline-serine-threonine phosphatase interacting protein 1, clone MGC:17132 IMAGE:4180398, mRNA, complete cds.
VERSION	BC008602
KEYWORDS	BC008602.1 GI:14250342
SOURCE	MGC.
ORGANISM	Homo sapiens.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE	1 (bases 1 to 1781)
JOURNAL	Strausberg,R. Direct Submission Submitted (25-MAY-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT	Contact: MGC help desk Email: cgabs-f@mail.nih.gov Tissue Procurement: David N. Louis, M.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: http://www.hgsc.bcm.tmc.edu/cdna/ Contact: villation@bcm.tmc.edu. Villation, D.K., Luna, R.A., Hale, S.M., Hullyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny,D.M., Gibbs,R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4100161.

FEATURES
source

Location/Qualifiers

1..1781
/organism="Homo sapiens"
/db_xref="LocusID:9051"
/db_xref="taxon:9606"
/clone="MGC:17132 IMAGE:4180398"
/tissue_type="Brain, anaplastic oligodendrogloma with
1p/19q loss"
/clone_id="NCI CGAP_Brn67"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
362..1612
/codon_start=1
/product="proline-serine-threonine phosphatase interacting
protein 1"
/protein_id="AAH08602.1"
/db_xref="GI:14250343"

CDS

BASE COUNT 395 a 525 c 572 g 289 t
ORIGIN
Alignment Scores:
Pred. No.: 4,786-133 Length: 1781
Score: 1913.50 Matches: 366
Percent Similarity: 92.79% Conservative: 20
Best Local Similarity: 87.98% Mismatches: 29
Query Match: 88.92% Indels: 1
Gaps: 1

US-09-068-377C-1 (1-415) x BC008602 (1-1781)

QY 1 MetMetAlaGlnLeuGlnPheArgSpAlaPheTyrPyrArgSpPheThrAlaHisThr 20
DB 362 ATGATGCCCGGAGCTGCAATTAAGATGCTTTGGTCAGGAGCTTCACACCCACACG 421
QY 21 G1YTYrGluValLeuLeuGlnArgLeuLeuAspGlyArgLysMetCysLysAspValGlu 40
DB 422 GGTACGAGGTGCTGCTGACGAGGCTTGTGATGCGAGGAAGATGCGAACAACATGAG 481
QY 41 G1uLeuLeuArgGlnArgAlaGlnAlaGluGluArgGlyGlyLysLysLysLysLys 60
DB 482 GAGCTACTGAGGAGGAGGCGGAGGCGGAGGCGGTACGGGAAGAGCTGTGTCAGATC 541
QY 61 AlaArgLysAlaGlyGlnGlnThrGlnMetLysSerLeuArgThrSerPheAspSerLeu 80
DB 542 GCACGAGAGGAGGAGGAGGAGGAGATCAACTCTGAGGAGCTCTCTGACTGCTTG 601
QY 81 LysGlnGlnThrGlnAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaLeuArgGlu 100
DB 602 AAGAGAGCAAAATGAGAAATGTCGAGCTCAGACATCCAGCTGCGCTGAG 661
QY 101 G1uLeuArgSerLeuGlnPheArgGluArgGlnLysGlnArgLysLysTyrglu 120
DB 662 GAGCTGCGGAGTCTCGAGAGATTCGTGAGAGGAGGAGGAGGAGGAAGATATGAG 721
QY 121 AlaIleMetAspArgValGlnLysSerLysSerLysSerLeuTyrLysLysThrMetLys 140
DB 722 GCCGTCATGAGACGGGCTCCAGAGAGCAAGCTGTGCTCTCAAGAGGCCCTGAGATC 781
QY 141 LysLysAlaTyrAspGlnLysCysArgAspAlaAspAlaGlnGlnAlaPheGluArg 160
DB 782 AAGAAGACATATGAGACAGATGCGCGGAGCGGAGCGAGCGAGCGCTTCAGCGC 841
QY 161 ValSerAlaAsnGlyHisGlnLysGlnValGluLysSerGlnAsnLysAlaLysGlnCys 180
DB 842 ATTAGCGGCAACGGCCACAGAGAGCTGAGAGAGTCAAGACCAAGCCAGCAGTGC 901

QY 181 LysGlnSerAlaThrGlnLysAlaGluArgValTyrArgGlnAsnIleGlnGlnLeuGluArg 200
DB 902 AAGAGATCGGCGACCGAGGAGGAGGAGGATATACAGGACAGACATTCGCGAGCGAGAG 961
QY 201 AlaArgThrGlnTyrPheGlnGlnLysArgThrThrCysGlnLysPheGlnLeuGlnGlu 220
DB 962 GTCCGGGCTGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1021
QY 221 PheAspArgLeuThrIleLeuArgAsnAlaLeuTyrValHisCysAsnGlnLeuSerMet 240
DB 1022 TTTGACCGGCTGACCATTCCTCCGCAAGCCCTGTGGTGCACAGCAACAGCTCCAGT 1081
QY 241 GlnCysValLysAspArgGlnLeuTyrGlnGluValArgLeuThrLeuGlnGluCysAsp 260
DB 1082 CAGTGTGTAGAGATATATAGCTCTACAGAGAGTGGCTGAGCGTGAAGGCTCCAGC 1141
QY 261 ValGlnGlyAspIleAsnGlyPheIleGlnSerLysSerThrGlyArgGluPropAla 280
DB 1142 ATAGAGCGCGACATGACAGCTTCATCCAGGCGCAAGAGCAGGAGGAGGAGGAGGAGG 1201
QY 281 ProValProTyrGlnAsnTyrTyrAspArgGluValThrProLeuIleGlySerProSer 300
DB 1202 CCGGTGCTCCACAGAACTATTACAGATCGGAGGAGTCAACCCGCTGACAGCAGCCTGC 1261
QY 301 IleGlnProSerCysGlyValIleLysArgPheSerGlyLeuLeuHisGlySerProLys 320
DB 1262 ATACAGCCCTGCTGCGGATGATTAAGAGCTTCTGAGACTGTGACGAGCAAGCTCCAG 1321
QY 321 ThrThrPro---SerAlaProAlaIleSerThrGlnThrLeuThrProThrProGluArg 339
DB 1322 ACCACTGCTTGGCAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1381
QY 340 AsnGluLeuValTyrAlaSerIleGluValGlnAlaThrGlnGlyAsnLeuAsnSerSer 359
DB 1382 AATGAGGAGTGTACACAGCCATCCGAGTGCAGAGATACAGAGGAACCCGCGCTCACA 1441
QY 360 AlaGlnAspTyrArgAlaLeuTyrAspTyrThrAlaGlnAsnSerAspGluLeuAspIle 379
DB 1442 GCCAGAGATACCGGGCGCTCTACGATTAATACGCCAGCAACCAATAGCTGAGCTG 1501
QY 380 SerAlaGlyAspIleLeuAlaValIleLeuGlnGlyGlnAspGlyTyrThrThrValGlu 399
DB 1502 TCCGGGAGACATCTGAGAGTGATCTGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1561
QY 400 ArgAsnGlyGlnArgGlyPheValProGlySerTyrLeuGlnLysLeu 415
DB 1562 AGGAACGGGAGCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1609
RESULT 5
HS094778 1656 bp mRNA linear PRI 05-JAN-1999
LOCUS
DEFINITION
Human PEST phosphatase interacting protein homolog (H-PiP) mRNA,
complete cds.
ACCESSION
U94778
VERSION
U94778.1 GI:4100161
KEYWORDS
SOURCE
ORGANISM
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 1656)
Wilson, L.A., Fields, D., Cruz, L., Lasky, L., Friesen, J. and
Siminovich, K.A.
TITLE
The human homologue of mouse PMP-PIP interactor protein
UNPUBLISHED
2 (bases 1 to 1656)
Wilson, L.A., Fields, D., Cruz, L., Lasky, L., Friesen, J. and
Siminovich, K.A.
AUTHORS
Direct Submission
Submitted (20-MAR-1997) Samuel Lunenfeld Research Institute, Mount
Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5,
Canada
JOURNAL

FEATURES

source 1..1656
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="activated leukocyte"

gene 1..1656
/gene="H-PIP"
217..1467
/gene="H-PIP"

CDS
/note="similar to Mus musculus PEST phosphatase interacting protein, encoded by GenBank accession Number U87814; encodes potential N-terminal coiled coil domain"
/codon_start=1
/product="PEST phosphatase interacting protein homolog"
/protein_id="AB00762.1"
/db_xref="GI:4100162"
/translation="MPOLQKFAKFCWCDFTAHQTEVILQRLDGRKCKMDLLR QRAAEERKELVQIARKAGQTEINSLRASFDSLKQOMENVSSHIQALATLEEL RSLKEFRERQKQRRKYEAVMDYVOKSKLSLYKAMEKSKTYEHKCLSDADDAFER ISANGHOKOVERQKARQCKDSATAEVYROSIAOLEKVAEWEQSHRTCEAFOL OEFRLTLIRNALVHNSNLSQCVKDELYEVLLEGGIDADISDFQAKSTGT EPPAPVPTONYDREVTPLTSSPGIOPSGDKRFSGLHSGPKTSLAASASTETL TPTPRNGVYTAIAVOTIGNPNASPADEYRAFYDYTANQNPEDLSAGDILEVLEGG BDGWTVERNGORGFVPSYLEKL"

misc_feature 430..490
/gene="H-PIP"
/note="encodes SH3 domain"

polyA_signal 1600..1605
/gene="H-PIP"

BASE COUNT 384 a 472 c 537 g 263 t

ORIGIN

Alignment Scores:

Pred. No.: 2,05e-130 Length: 1656
Score: 1877.50 Matches: 358
Percent Similarity: 91.59% Conservative: 23
Best Local Similarity: 86.06% Mismatches: 34
Query Match: 87.24% Indels: 1
Gaps: 1

US-09-068-377c-1 (1-415) x HSU944778 (1-1656)

QY 1 MetMetAlaGlnLeuGlnPheArgAspAlaPheTrpCysArgAspPheThrAlaHisThr 20
DB 217 ATGATGCCCCACAGCTCCAGTCAAAAGATGCTTTGTGTGTCAGAGCACTTACAGCCACAG 276

QY 21 GtYTrGluValLeuLeuGlnArgPheLeuAspGlyArgGlySerMetCysLysAspValGlu 40
DB 277 GGCTACGAGGTGCTCTCTGAGCGGCTTGTGATGGCAGAGAGATGTGCAAAACATGGAG 336

QY 41 GtLeuLeuArgGlnArgAlaGlnAlaGluArgGlyArgGlyLysGluLeuValGlnIle 60
DB 337 GAGCTACGTGAGGCACAGAGGCCACGCGGAGGCGGTACGGGAAGAGCTGTGTCAGATC 396

QY 61 AlaArgLysAlaGlyGlyGlnThrGluMetAsnSerLeuArgTrpSerPheAspSerLeu 80
DB 397 GCACGGAAGGACGAGTGGCCAGACGAGATCACTCCGAGAGGCTCTCTTGACTCCTTG 456

QY 81 LysGlnGlnThrGluAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaLeuArgGlu 100
DB 457 AAGCAGCAAAATGGAGATGTGGCCAGCTCACACATCCAGTGGCCCTGACCTGCGTAG 516

QY 101 GtLeuArgSerLeuGluGluPheArgGluArgGlnLysGluGlnArgLysLysTrpLeu 120
DB 517 GAGCTGGCGAGTCTCAAGGAGTTTGTGAGAGCGCAAGAGCAAGAGAGATATAG 576

QY 121 AlaIleMetAspArgValGlnLysSerLysLeuSerLeuTrpLysLysThrMetLysLeu 140
DB 577 GCGCTGATGACCGGGTCCAGAGAGCAAGCTGTCTCTACAGAAAGCATGAGATCC 636

QY 141 LysLysAlaTrpArgGlnLysCysArgAspAlaAspAspAlaGluGlnAlaPheGluArg 160
DB 637 AAGAGACTTACGACATAGTCTGAGTCTGGAGCGAGCGGATCCAGCTTCCAGGCG 696

QY 161 ValSerAlaAsnGlyHisGlnLysGlnValGluLysSerGlnAsnLysAlaLysGlnCys 180
DB 697 ATTACCGGCACAGCGCCACAGAGAGAGTGGAGAAGATGACGAACAAAGCCAGCAGTGC 756

QY 181 LysGluSerAlaThrGluAlaGluArgValTrpArgGlnAsnIleGluGlnLeuGluArg 200
DB 757 AAGGACTGCGCCACAGCAGAGAGCGGATATACAGCGAGACATTTGACCTGAGAGAG 816

QY 201 AlaArgThrGlnTrpGluGlnGlnHisArgThrTrpCysGluAlaPheGlnLeuGlnGlu 220
DB 817 GTCCGGGTGAGTGGAGACAGAGACCGGACACCTGTGAGGCTTTCACCTCCAAAG 876

QY 221 PheAspArgLeuThrIleLeuArgAsnAlaLeuTrpValHisCysAsnGlnLeuSerMet 240
DB 877 TTGACCGCGGTGACCATTTCTCCGACAGCCCTGTGGTGTCACAGACACACTCTCCATG 936

QY 241 GlnCysValLysAspAspGluLeuTrpGluGluValArgLeuThrLeuGluGlyCysAsp 260
DB 937 CAGTGTGCAAGATGATGACTCTACAGAGAGTGGCGGTGACGCTGGAAGCTGCAGC 996

QY 261 ValGluGluAspIleAsnGlyPheIleGlnSerLysSerThrGlyArgGluProProAla 280
DB 997 ATGACGCCGACATCGACAGTTTCAATCCAGGCCAAGACACGGGACAGACCCCGCT 1056

QY 281 ProValProTyrGlnAsnTrpTrpAspArgGluValThrProLeuIleGlySerProSer 300
DB 1057 CCGGTGCGCTACACAGACTATTAGATGGAGGACACCCGCTGACAGACGCGCTGCGC 1116

QY 301 IleGlnProSerCysGlyValIleLysArgPheSerGlyLeuLeuHisGlySerProLys 320
DB 1117 ATACAGCCGCTCGTGGGATGATAAGAGTTCCTGTGACTGTGCACAGGAATGCCAAG 1176

QY 321 ThrTrpPro---SerAlaProAlaAlaSerThrGlnThrLeuThrProThrProGluArg 339
DB 1177 ACCACTGCTTGGCAGCTTCTGTGCTGCGTCGTCACAGAGACCTGTACCCCGCAGAGG 1236

QY 340 AsnGluLeuValTrpAlaSerIleGluValGlnAlaThrGlnGlyAsnLeuAsnSerSer 359
DB 1237 AATGAGGCTGTCTACACAGCATGCAATGAGAGGATACAGGGAACCCGCGCTCACCA 1296

QY 360 AlaGlnAspTrpArgAlaLeuTrpAspTrpTrpAlaGlnAsnSerAspGluLeuAspIle 379
DB 1297 GCGCAGAGATACCGGCGCTTCTACGATTATACAGCGCAGAACCCAGACTGTGACCTG 1356

QY 380 SerIleGluAspIleLeuAlaValIleLeuGluGlyGluAspGlyTrpTrpThrValGlu 399
DB 1357 TCCGGGGAGACATCTCGAGGTATCTGGAAGGGGAGATGCTGTGTGAGTGTGAG 1416

QY 400 ArgAsnGlyGlnArgGlyPheValProGlySerTrpLeuGluLysLeu 415
DB 1417 AGGAACGGGACAGCGTGGCTGTCTGCTTCCATCCAGGAGAGCTT 1464

RESULT 6

AF038602 1382 bp mRNA linear PRI 02-FEB-1999

LOCUS Homo sapiens CD2 binding protein 1 short form mRNA, complete cds.

DEFINITION AF038602

ACCESSION AF038602

VERSION AF038602.1 GI:2921548

KEYWORDS

SOURCE

ORGANISM

Homo sapiens.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1382)

AUTHORS L.J., Nishizawa, K., An, W., Hussey, R.E., Lialios, F.E., Salgia, R., Sunder-Plassmann, R. and Reinherz, E.L.

TITLE A cdcs-1-like adaptor protein (CD2BP1) interacts with the CD2 cytoplasmic domain and regulates CD2-triggered adhesion

JOURNAL EMBO J. 17 (24), 7320-7336 (1998)

MEDLINE 99077800

PUBMED 9857189

REFERENCE 2 (bases 1 to 1382)

AUTHORS

L.J., Nishizawa, K., An, W., Hussey, R.E., Lialios, F.E.,
Sunder-Plassman, R. and Reinherz, E.L.

TITLE

Direct Submission

JOURNAL

Submitted (15-DEC-1997) Immunobiology, Dana-Farber Cancer
Institute, 44 Binney Street, Boston, MA 02115, USA

FEATURES

Location/Qualifiers

CDS

```

1..1382
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="T cell"
1..1194
/note="CD2AP15"
/codon_start=1
/product="CD2 binding protein 1 short form"
/protein_id="A011958.1"
/db_xref="GI:2921549"
/translation="MFPOLQKDAFWCRDFTAHNGEYLLQRLIDGRMKMDMELL
ORADBERYGELEVOIARKAGOTEINSLRSPSLKQOMENVSHTOLATREEL
RSLEPERROKRYEAVMDRVOKSLSYRKAMEKRYEOKCRADADAQAFER
ISANGHOKOYKESONKAROCKSDATEAEVRYROSLAOLEKRAEMOEHRTEAFOL
OEPRRLTILRNALMVSNOISMOCVKDELEEVRLTEGSDIADIDSFQAKSTGT
EPREVEVLADSAASRFSGLIHSGPKTSLASASTETLPTPERNEGVYALAVOEI
QGNPASPAQERFALDYTAQNPDELDSAGDILEVILEGEDMWTVERNGRFPVGS
YLEKL"

```

BASE COUNT 344 a 375 c 453 g 210 t

ORIGIN

Alignment Scores:

Pred. No.: 6.65e-121 Length: 1382
Score: 1748.00 Matches: 341
Percent Similarity: 86.78% Conservative: 20
Best Local Similarity: 81.97% Mismatches: 35
Query Match: 81.23% Indels: 20
DB: 9 Gaps: 2

US-09-068-377c-1 (1-415) x AF038602 (1-1382)

```

QY 1 MetMetAlaGlnLeuGlnPheArgAspAlaPheTrpCysArgAspPheThrAlaHisThr 20
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 ATGATGCCCGCCAGCGCATTCACAAATGCCCTTTGGTGCAGGACTTCACAGCCACACG 60
QY 21 GlyTyrGluValLeuLeuGlnArgLeuLeuAspGlyArgLysMetCysLysAspValGlu 40
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 GCGTACAGAGTGGCTGTGCAGCGGCTTCTGTGATGGCAGAGATGTGCAAGACATGAG 120
QY 41 GlnLeuLeuArgGlnArgAlaGlnAlaGlnGluArgTyrGlyLysGlnLeuValGlnIle 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 GAGCTACTGAGGCGAGGCGCCAGCGAGGAGGCGGTACGGGAAGGAGCTGTGCAGATC 180
QY 61 AlaArgLysAlaGlyGlnThrGlnMetArgSerLeuArgThrSerPheAspSerLeu 80
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 GCACGGAAGGAGGAGTGGCCAGACGAGATCACTCCCTGAGGGCTCTCTTGACTCCTTG 240
QY 81 LysGlnGlnThrGlnAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaLeuArgLys 100
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 AAGCAAGCAATGCGATGTGGCAGCTTCACACATCCAGCTGGCCCTGACCTGGGTAG 300
QY 101 GlnLeuArgSerLeuGlnGluPheArgGluArgGlnLysGlnGlnArgLysLysTyrGlu 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 GAGTGGGAGGTCTCGAGGAGTTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 360
QY 121 AlaIleMetAspArgValGlnLysSerLysLeuSerLeuTyrLysLysThrMetGlnSer 140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 GCGGTATGAGCGCGGTCCAGAGACAGCAAGCTGTCCCTTCACAGAGGAGGAGGAGGAG 420
QY 141 LysLysAlaTyrAspGlnLysCysArgAspAlaAspAspAlaGlnGlnAlaPheGlnArg 160
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 AAGAAAGCATACGACGAGAGTCCGCGGACCGGACGACGCGGAGGAGGAGGAGGAGG 480
QY 161 ValSerAlaAsnGlnLysGlnLysGlnValGlnLysSerGlnAsnLysAlaLysGlnCys 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 481 ATTAGCCCAACGGCCACGACGAGACGAGTGGAGAGAGTACGAACAAGCCAGCGAGTGC 540

```

```

QY 181 LysGlnSerAlaThrGlnAlaGlnArgValTyrArgGlnAsnIleGlnGlnLeuArg 200
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 541 AAGAGCTGCGCCAGCGAGGACAGCGGATATACAGGACAGACATTCGCGAGTGGAGAG 600
QY 201 AlaArgThrGlnTyrPgluGlnGlnHisArgThrThrCysGlnAlaPheGlnLeuGlnGlu 220
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 601 GTCCGCGCTGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 660
QY 221 PheAspArgLeuThrIleLeuArgAsnAlaLeuTyrPvalHisCysAsnGlnLeuSerMet 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 661 TTTGACCGCGCTGACATTCCTCCGCAACGCCCTGTGGTGCACAGCAACGAGCTTCACAG 720
QY 241 GlnCysValLysAspAspGlnLeuTyrGlnGlnValArgLeuThrLeuGlnGlnCysAsp 260
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 721 CAGTGTGTCAAGGATGATGAGTCTTACAGGAGAGTGGCTTACGCTGGAAAGGCTGCAC 780
QY 261 ValGlnGlyAspIleAsnGlnPheIleGlnSerLysSerThrGlyArgGlnProGlnAla 280
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 781 ATAGACGCGGACATGACATGAGTTTCAATCCAGCCAGAGAGGAGGAGGAGGAGGAGGAG 840
QY 281 ProValProTyrGlnAsnTyrTyrAspArgGlnValThrProLeuIleGlySerProSer 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 841 GAGGTCCGCGCTTGCAGAC----- 858
QY 301 IleGlnProSerCysGlyValIleLysArgPheSerGlyLeuLeuHisGlySerProLys 320
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 859 -----ACGCGACCTCTAGGTTCTCTGAGCTGCTGACGAGTCCAG 903
QY 321 ThrThrPro---SerAlaProAlaAlaSerThrGlnThrLeuThrProThrProGluArg 339
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 904 ACCACTGTGTGGCAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 963
QY 340 AsnGlnLeuValTyrAlaSerIleGlnValGlnAlaThrGlnGlyAsnLeuAsnSerSer 359
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 964 AATGAGGCTGTCTTACACAGCATGCGAGTGCAGGAGATACAGGAGAAACCGGCTCCACCA 1023
QY 360 AlaGlnAspTyrArgAlaLeuTyrAspTyrThrAlaGlnAspSerAspGlnLeuAspIle 379
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1024 GCCCAGAGATACCGGGGCGCTCTACGATTTATACAGGCGAGAACCCAGATAGCTGACCTG 1083
QY 380 SerAlaGlyAspIleLeuAlaValIleLeuGlnGlyGlnAspGlyTyrThrThrValGlu 399
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1084 TCCCGCGGAGACATCTCTGAGGTGATCTCGAAGGAGGAGGAGGAGGAGGAGGAGGAG 1143
QY 400 ArgAsnGlnArgGlyPheValProGlySerTyrLeuGlnLysLeu 415
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1144 AGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1191

```

RESULT 7
BC002123
LOCUS
DEFINITION
Mus musculus, proline-serine-threonine phosphatase-interacting protein 2, clone MGC:6634 IMAGE:3494357, mRNA, complete cds.
ACCESSION
BC002123
VERSION
BC002123.1 GI:12805312
KEYWORDS
MGC.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Mus.
REFERENCE
1 (bases 1 to 1498)
AUTHORS
Strausberg, R.
JOURNAL
Submitted (31-JAN-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK
Contact: MGC help desk
Email: cga@nci.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcdelpax11.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/MLN at: <http://image.llnl.gov>
 Series: IRAP Plate: 8 Row: k Column: 8.
 Location/Qualifiers
 1.1498

FEATURES

source

/organism="Mus musculus"
 /db_xref="taxon:10090"
 /map="CECH II"
 /clone="MGC:6634 IMAGE:3494357"
 /tissue_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMV."
 /clone_id="NCI_CGAP_Lu29"
 /lab_host="DH10B"
 /note="Vector: PCMV-SPORT6"
 150.1154
 /codon_start=1
 /product="proline-serine-threonine phosphatase-interacting protein 2"
 /protein_id="AA02123.1"
 /db_xref="GI:12805313"
 /db_xref="LocusID:19201"
 /translation="MTGSLFKGNFSDILSTIGYDSTIOHLNKRKNCKEPEFLAKE RASIEEYKGLDLNLSRKKPQSEINTLRALREKQOVDAQCHIDLAQTLREBA RKMEPERKOKLRKTETIMDAHKRNQFKKAMDAKRYNOKRDEDAQVRA SANVANOROEKLFVLAITSKTAVEDSDKAYMLHIMLEVRDMSSEHAKCEVEA QCEERINFEKNALMLHLNLSQCVANDEYEVKRSLEKSIENKIOYEVNOKRTOO TPAPPIYENFYSFORNAAPPEKTPGNRGRPLPVKRIPIDDPYSVEDYSLIXO"

CDS

BASE COUNT 442 a 358 c 366 g 332 t
 ORIGIN

Alignment Scores:

Pred. No.: 7.18e-46 Length: 1498
 Score: 737.50 Matches: 146
 Percent Similarity: 58.52% Conservative: 67
 Best Local Similarity: 40.11% Mismatches: 112
 Query Match: 34.27% Indels: 39
 DB: 10 Gaps: 5

US-09-068-377C-1 (1-415) x BC002123 (1-1498)

QY 7 PheATGAsPAlaPheTTPCysArGAsPheThAlaHisThcGluValLeuLeu 26
 Db 165 TTCAGGGGAACTTCTGGAGCAGGACATTCACACCATTCGGCTACGACGCAATTATC 224
 QY 27 GlnATGLeuLeuAspGlyArgGlyMetCysLysAspValGluGluLeuArgGlnArg 46
 Db 225 GACCACTGTAACAATGGCCGTAAAGAACTGCAAGAGGTTGAAGACTTTTGAAGAAAGA 284
 QY 47 AlaGlnAlaGluGluArgGlyArgGlyLysGluLeuValGlnIleAlaArgLysAlaGly--- 65
 Db 285 GCATCAATTGAAGAAATATGGCAAGAACTGCTGAACCTCTCCAGGAAGAAGCCGTGT 344
 QY 66 GlyGlnThrGluMetAsnSerLeuArgThrSerPheAspSerLeuLysGlnGlnThrGlu 85
 Db 345 GGACAGTCTGAGATCAATACCTTGAAGAGCCCTTGAGCTTTCACAGCACAAGTGAAC 404
 QY 86 AsnValIGlySerAlaHisIleGlnLeuAlaLeuAlaLeuArgGluGluLeuArgSerLeu 105
 Db 405 AACGTGGCTCAATGTCACTTCACTTCCGAGACTTAAGAGAAAGCCAGAGAGATG 464
 QY 106 GlnGluPheArgGluArgGlnLysGluGlnArgLysLysTyrgLysAlaIleMetAspArg 125
 Db 465 GAACAATTCGCGGAAGACAAAGCTGCAGCGGAAGAAAGACAGAAATATGATGCT 524

QY 126 ValGlnLysSerLysLeuSerLeuThrLysLysThcMetLysSerLysAlaIleArg 145
 Db 525 GCCACAAAGCAAGAAAGACGACACTTCAAGAAAGCCATGGATGCCAAGAAATATGAG 584
 QY 146 GlnLysCysArgAspAlaAspAspAlaGluGlnAlaPheGluArgValSerAlaAsnGly 165
 Db 585 CAAAGAGCCGCGGCAAGATGAGGACAGACCAAGCTGTCACCGCAGCCCAATGCGCC 644
 QY 166 HisGlnLysGlnAlaGlnLysSerGlnAsnLysAlaLysGlnCysLysLysSerAlaThr 185
 Db 645 AACCAAGCGCAACAGAAAGCTTTTGTGAACGTGCCCACTTCAAGACTGCAGTAGAG 704
 QY 186 GlnAlaGluArgValTyArgGlnAsnIleGluGlnLeuLysGluArgAlaArgThrGlu 205
 Db 705 GATTTCAGACAAACGATGATGTTGACATTCACATTCATCTGAGAGTTGACAAAGCTGG 764
 QY 206 GlnGlnGlnHisArgThrThrCysGlnAlaPheGlnLeuGlnGluPheAspArgLeuThr 225
 Db 765 CAGAGTGAACCACTTAAGCCCTGCGAGGTGTCGAGGCTCAGAGATGTAACGATCAAC 824
 QY 226 IleLeuArgAsnAlaLeuThrPValHisCysAsnGlnLeuSerMetGlnCysValLysAsp 245
 Db 825 TTCTTCGGAATGCACTGTGTGATCTGATTCAGCTGTCTCAACAATGTGTGCAAT 884
 QY 246 AspGluLeuThrGluGluValArgLeuThrLeuGluGlyCysAspValGluGlyAspIle 265
 Db 885 GACGAGATGATGAACAACCTCCGTAAAGATTGAAGACGTGACGATTCAGAGACATC 944
 QY 266 AsnGlyPheIleGlnSerLysSerThrGlyArgGluProPAlaProValProTyrGln 285
 Db 945 CAGTATTTGTGAACACGAGGAACTGACAGCTCCCTCCAGACCCATCATATGATGAG 1004
 QY 286 AsnTyTYrAspArgGluValThrProLeuIleGlySerProSerIleGlnProSerCys 305
 Db 1005 AACTTCTAC-----TCTCTCTAG-----1022
 QY 306 GlyValIleLysArgPheSerGlyLeuLeuHisGlySerProLysThrProSerAla 325
 Db 1023 -----CGGATCTGCGCCGCCAGAAAGACTACAGGC 1055
 QY 326 ProAlaAlaSerThrGluThrLeuThrProThrProGluArg-----AsnGluLeuVal 343
 Db 1056 CCCAATCCAGCAAGAGAGACACTCTCCCTGCTCCCTTAAGGAGATACAGATGACCTGAC 1115
 QY 344 TyrAlaSerIleGluValGlnAlaThrGlnGlyAsnLeuAsnSerSerAlaGlnAspTyr 363
 Db 1116 TACTCTGTGTTGAA-----GATTAC 1136
 QY 364 ArgAlaLeuTyr 367
 Db 1137 AGTTGCTCTAT 1148
 RESULT 8
 LOCUS MWU18101
 DEFINITION Mus musculus mRNA for macrophage
 actin-associated-tyrosine-phosphorylated protein.
 VERSION Y18101.1 GI:3947711
 KEYWORDS macrophage actin-associated-tyrosine-phosphorylated protein; mayp
 gene.
 SOURCE
 ORGANISM Mus musculus.
 Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Yung, T.G., Soldera, S. and Stanley, E.R.
 A novel macrophage actin-associated protein (MAYP) is
 tyrosine-phosphorylated following colony stimulating factor-1
 stimulation
 JOURNAL J. Biol. Chem. 273 (46), 30638-30642 (1998)
 MEDLINE 99023997
 PUBMED 9804836

REFERENCE 2 (bases 1 to 1005)
AUTHORS Soldera, S.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-1998) S. Soldera, Albert Einstein College of Medicine, 1300, Morris Park Avenue, Bronx, New York 10461, USA
FEATURES
source
1..1005
/organism="Mus musculus"
/strain="(BALB/c x A.CA)F1"
/db_xref="taxon:10090"
/cell_line="BAC1.2F5"
1..1005
/gene="mayp"
1..1005
/gene="mayp"
/codon_start=1
/product="macrophage
actin-associated-tyrosine-phosphorylated protein"
/protein_id="CAA77027.1"
/db_xref="GI:3947712"
/translation="MTSLFGKEMWSTDLISITIGYDIIIOHNGRNCKEPEFLKE
RSTIEKYGKDLNLSRKKPGQSEINILKRALVEFKOVDNACHOLATLKEEA
RKMEFRKOKLORRKTETIMDAHKORNAOFKAMADAKNETEQCRDKEQAVHR
SANYANQOQELVKLATSKTAVEDSKAYMLNMLEKVEDQSEHAKCEVFEA
QECERINFERNALMLHLNQLSQOCVANDEMEQVRKSLSETCIEDIDQYFNQRTGQ
TPPAPIYENFSPQRNAPGPKTGPBARPGPLVPKRIIPDDPDYSVEDYSILYO
"

BASE COUNT 325 a 227 c 259 g 194 t
ORIGIN

Alignment Scores:
Pred. No.: 5,16e-46 Length: 1005
Score: 736.50 Matches: 146
Percent Similarity: 58.52% Conservative: 67
Best Local Similarity: 40.11% Mismatches: 112
Query Match: 34.22% Indels: 39
DB: 10 Gaps: 5

US-09-068-377C-1 (1-415) x MMU8101 (1-1005)

OY 7 PheargaspAlaphetrcPysargaspPhehralahisthGlytyrclValleuLeu 26
Db 16 TTCAAGGGGAACTTCTGGACGACGACATTCACACACCATTCGACGACATTCAC 75
OY 27 GlnargleuLeuAspGlyArgtylMetCysLysPvalGlnGluLeuLeuArgGln 46
Db 76 CAGCACCTCGAACAAAGCCGTAAGACCTGCAAGAGAGTGTGTAAGAGAGAGA 135
OY 47 AlaGlnaLagLugLuarGtyrclLysgluLeuValGlnlleAlaArgLyslaagly 65
Db 136 GCATCAATTGAAGAAATATGCGCAAGACCTGCTGCAACCTCTCCAGGAAGACCGTGT 195
OY 66 GlyGlnthrGluMetAsnSerleuArgthrSerPheasSerleuLysGlnGlnthrGlu 85
Db 196 GCACAGCTCTGAGATCAATACCTTGAAGAGAGCCCTTGAGTCTTCAGACGACAGTAGAC 255
OY 86 AsnArgLysSerAlaHisileGlnleuAlaLeuAlaLeuArgGlnGluLeuArgSerleu 105
Db 256 AACGCGCTCAATGTCACATTCAGCTTGGCGACGACTCTAAGAGAGAGCCGCAAGAGAG 315
OY 106 GlnGluPheargGluArgGlnLysGlnGlnArgLysLysTyrgLualaIleMetaspArg 125
Db 316 GAAGAAATCCCGGAAAGACGAGCTGCGAGAAAGAGAGAAACATATATGATGCT 375
OY 126 ValGlnLysSerLysLeuSerleuTyrlLysLysThrMetGlnSerLysLysAlaTyraP 145
Db 376 GCCCAAGCAAAAGGACGACAGTTCAGAAAGACGATGATGCCAAGAGATTAATGAG 435
OY 146 GlnLysCysArgaspAlaAspAlaGlnGlnAlaIlePheGlnArgValSerlaAsngly 165
Db 436 CAAGAAGTCCGCGACAAAGATGAGCAGAGCAAGCTGTCCACGCGACCCCAATGTGGCC 495
OY 166 HisGlnLysGlnValGlnLysSerGlnAsnLysAlaLysGlnCysLysGlnSerAlaThr 185

Db 496 AACCAAGCGCAACAGAAAAGCTTTTGTGAACCTGGCCACTTCAAAGACTGCGAGTAGAG 555
OY 186 GlnAlaGlnArgValIlyrArgGlnAsnIleGlnGlnLysGlnArgAlaArgThrGluTrp 205
Db 556 GATTTCAGACAAACCGTATGATGTTGACATCATCATCATCATCATCATCATCATCATCAT 615
OY 206 GlnGlnGlnHisArgThrThrCysGlnAlaPheGlnLeuGlnGlnIlePheAspArgLeuThr 225
Db 616 CAGAGTGAACACATTAAAGCCCTCGAGAGCTGTGAGGCTGAGAGATGTGACAGATCAAC 675
OY 226 IleLeuArgAsnAlaLeuTrpValHisCysAsnGlnLeuSerMetGlnCysValLysasp 245
Db 676 TTCTCCCGGAAATCAGCTGTGTTGATCATCATCATCATCATCATCATCATCATCATCAT 735
OY 246 AspGlnLeuLeuTyrgLlnGlnLualaArgLeuThrLeuGlnGlnLysCysAspValGlnGlyAspIle 265
Db 736 GAGCAGATGTATGAGCAAGTCCGTAAGACATTGAAACATGACATGACATGACATGACATC 795
OY 266 AsnGlyPheIleGlnSerLysSerThrGlyArgGlnProAlaProAlaProAlaProTyrgLln 285
Db 796 CAGTATTTTGTGAAACGACGAGGAAACCTGACAGACTCCCGCAGACCCATCATGATGATGAG 855
OY 286 AsnTyrlTyraSpArgGlnValThrProLeuIleGlnLysSerProSerIleGlnProSerCys 305
Db 856 AACTCTTAC-----TCCTCTCAG-----873
OY 306 GlyValIleLysArgPheSerGlyLeuLeuHisGlySerProLysThrThrProSerAla 325
Db 874 -----CGGAATGCTGCCCCACGACGACGAAAGACTACAGGG 906
OY 326 ProAlaLaserThrGlnThrLeuThrProThrProGlnArg-----AsnGlnLeuVal 343
Db 907 CCCAATCCAGACAGAGAGAGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 966
OY 344 TyrlAserIleGlnValGlnAlaThrGlnGlnGlyAsnLeuAsnSerSerAlaGlnAspTyrl 363
Db 967 TACTCTGTGCTGCA-----GATTAC 987
OY 364 ArgAlaLeuTyrl 367
Db 988 ACTTGTCTCTAT 999

RESULT 9
AX058229 1580 bp DNA linear PAT 17-JAN-2001
LOCUS AX058229
DEFINITION Sequence 99 from Patent WO0077040.
ACCESSION AX058229
VERSION AX058229.1 GI:12310730
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1580)
AUTHORS Yue, H., Tang, Y.T., Hillman, J.L., Lal, P., Bandman, O., Baughn, M. R.,
Azimzai, Y., Yang, J., Reddy, R. and Lu, D. A.
TITLE Human intracellular signaling molecules
JOURNAL Patent: WO 0077040-A 99 21-DEC-2000;
Incyte Genomics, Inc. (US)
FEATURES
source
1..1580
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 498 a 352 c 372 g 358 t
ORIGIN

Alignment Scores:
Pred. No.: 9.92e-46 Length: 1580
Score: 736.00 Matches: 144
Percent Similarity: 63.80% Conservative: 64
Best Local Similarity: 44.17% Mismatches: 112
Query Match: 34.20% Indels: 6

DB: 6 Gaps: 3

US-09-068-377c-1 (1-415) x AX058229 (1-1580)

QY 2 MetAlaGlnLeuGlnPheArgAspAlaPheTrpCysArgAspPheThrAlaHisThrGly 21

DB 62 AAGAGCGCGCTCAGCTGTCAAGGAAACCTTTGGAGTGCAGCATCCTCGACCATCGCG 121

QY 22 TTTGluValLeuGlnLeuArgLeuLeuAspGlyArgGlyLysMeCysLysAspAlaGlu 41

DB 122 TATGACACATATTATCCACATCTGAACATGCGCCGACAGAACTGCAGAAAGATTGACAGC 181

QY 42 LeuLeuArgGlnArgAlaGlnAlaGluArgGlyArgGlyLysGluLeuValGlnIleAla 61

DB 182 TTTCTAAAGAAAGAGGACCAATTGACAGAGGTATGCAAGATCTGCTCAACTCTCT 241

QY 62 ArgGlyAlaGly---GlyGlnThrGluMetAsnSerLeuArgTrpSerPheAspSerLeu 80

DB 242 AGGAAGAACCCGTGTGGACAGTCTGAATCAACACCCCTGAACCGGCGCTTGAAGTCTTC 301

QY 81 LysGlnGlnThrGluAsnValGlySerAlaHisIleGlnLeuAlaLeuLeuArgGlu 100

DB 302 AAGCAGCAAGTAGACATGTGGCAGCATGTCCATTCAGCTTGCACAGAGTTAAGAGA 361

QY 101 GtuleuArgSerLeuGluGluPheArgGluArgGlnLysGluGlnArgLysLysTrpGlu 120

DB 362 GAGCCGAGAGATGAGAAATTCAGGCGAAAGCAAAACTACAAAGCAAAAGAACAGACAG 421

QY 121 AlaIleMetAspArgValGlnLysSerLysLeuSerLeuTrpLysThrMetGluSer 140

DB 422 CTCATTAATGATGCTATCATTAACAAACAAAGACCTTCAATTAAGAAACCAATGATGA 481

QY 141 LysLysAlaTrpAspGlnLysCysArgAspAlaAspAlaGluGlnAlaPheGluArg 160

DB 482 AAGAAAGAACTATGACAGAGAAATGCGGAGCAAAAGATGAGGAGCAAGAGCCCTGACCGG 541

QY 161 ValSerAlaAsnGlyHisGlnLysGlnValGluLysSerGlnAsnLysAlaLysGlnCys 180

DB 542 ACTGCCAACCTGTGTACACCGAAGCAACAAAGAAAGCTTTTGTGAATGTGCAACTTCA 601

QY 181 LysGluSerAlaThrGluAlaGluArgValTrpArgGlnAsnIleGluGlnLeuGluArg 200

DB 602 AAGACCGCGAGTAGAGACTCAGACAAAGCATTCATCTGCACATGGCACCTGGATAG 661

QY 201 AlaArgTrpGluTrpGluGlnLysHisArgTrpThrCysGluAlaPheGlnLeuGlnGlu 220

DB 662 GTCCGAGAAAGTAGGACAGTAGCAGCATCAGAGCCCTGCGAGCATTTGAGGCTCAGAGA 721

QY 221 PheAspArgLeuThrIleLeuArgAsnAlaLeuTrpValHisCysAsnGlnLeuSerMet 240

DB 722 TGTGACAGCAATAACTTCTTCGGAAATGATGTGTGATGATGTAATCACTGTCACAA 781

QY 241 GlnCysValLysAspAspGluLeuLeuTrpGluGlnValArgLeuThrLeuGlnLysAsp 260

DB 782 CAATGTGTCCACAGATGATGAATGTACGAAAGAGTCCGAAAGAGTTTAAAGATGTGCACG 841

QY 261 ValGluGlyAspIleAsnGlyPheIleGlnSerLysSerThrGlyArgGluProAla 280

DB 842 ATTCAGAGGACATGATGATCTTGTGAATCAACGCAAAACTGACAGATTCACACAGCA 901

QY 281 ProValProTyrGlnAsnTrpTyrAspArgGluValIleProLeuIleGlySerProSer 300

DB 902 CCCATCATCATGATGAATTTCTACCTCCCGAAGAAATGCGCCAGCAGGAAGAGCT 961

QY 301 IleGlnProSerCysGlyValIleLysArgPheSerGlyLeuLeuHisGlySerProLys 320

DB 962 ACAGGCGCTAAC-----TTGGCAAGAGAGAGACCCCTC-----CCAATTCTTAAA 1006

QY 321 ThrThrProSerAlaPro 326

DB 1007 AGCTCACACAGATGATCCC 1024

RESULT 10

AC116689/c
LOCUS
DEFINITION
AC116689
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

AC116689
Mus musculus clone RP24-201N16, WORKING DRAFT SEQUENCE, 12 ordered
pieces.
AC116689.3 GI:21622817
HTG: HTGS_PHA5E2; HTGS_DRAFT; HTGS_FULLTOP.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 161469)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus, clone RP24-201N16
Unpublished
2 (bases 1 to 161469)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
Anderson,S., Batta,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Bouhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collamore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Lacroque,K., Lamazares,R.,
Lander,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McKernan,K., Meldrum,J., Menus,L.,
Mihova,T., Mienge,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,J., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollard,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (02-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 161469)
Birren,B., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Batta,N., Bastien,V., Bloom,T., Boguslavsky,L., Bouhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Lander,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrum,J., Menus,L., Mihova,T., Mienge,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,J., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission

TITLE
JOURNAL
COMMENT

Submitted (28-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 28, 2002 this sequence version replaced gi:21592236.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information

[illegible]

QY 279 ----- 279
Db 136630 AGGTTTCCTGTCGTCTAGCCCTGGCTGCTCTGGCACTCATTGTAGACCAAGCTGGCC 136571
QY 279 ----- 279
Db 136570 TCGAACTCAGAAATCCGCTGCTGCTGCCGAGTGTGGATTAAAGCGTGGCCA 136511
QY 279 ----- 279
Db 136510 CCAGCCCGGCTAAACAAATCTTTAAAAAAGCCAGCAGCTGTTCATGC 136451
QY 279 ----- 279
Db 136450 CTTTAATCCCTCATGGGGTGGCAGAGCGAGTGCCTCTGACAGAGCTCTCGCTGCT 136391
QY 280 ----- -AlaProValProTyrGlnAsnTyrTyrAspArg 290
Db 136390 CTTGGCTCATGCAATTTGCTCTGTGAGCTCCGCTCTTATCAGAACTACTAGACAG 136331
QY 291 GlnValThrProLeuIleGlySerProSerIleGlnProSerCysGlyValIleLys --- 309
Db 136330 GAGGTGACCCCACTGATTGGCAGCCCTAGCTCCAGCCCTCTGTGTGTGAT -AAAGAG 136272
QY 309 ----- 309
Db 136271 GTGAGCACCCGATGAACCTGATATGATCAGACACAGCCGGGGTTCAGCCTGACTTCCA 136212
QY 309 ----- 309
Db 136211 CATTCATGCTGTAAGAGCTCCTGTGGTGTGTGTACCTGTTCAGAAAGTGGCATCCA 136152
QY 309 ----- 309
Db 136151 TTGAGATAGTCCAGATCAAACTACCCTGAGAGCCCGGGTCCAGAGCATCAAGTA 136092
QY 309 ----- 309
Db 136091 CTAAATTACTCACACTCCAGTTATTTCACCAAGCGCTATGGGGAGAGTCCCGGGCAACTGCC 136032
QY 309 ----- 309
Db 136031 CACTCTGCTCTCTGTCTGTCTGTGCGCTACTAGAGCGAGGTAGTGTCTTAACTGAGA 135972
QY 309 ----- 309
Db 135971 GGTAGCACGGGTGACTTCCAGGTGGCTTATGAATGGCCCAAGCCAGACATTGGATTTC 135912
QY 309 ----- 309
Db 135911 ACCAGGCTTTTAGAGGGCCATCTCCCTGGCTGGGTGTGTAGTCAAGTCCCAAGCCCTAGT 135852
QY 309 ----- 309
Db 135851 TCTGTGCTCTGTGGTGAAGCTGAGTCAATTTGGCAGCCCAAGAAAAGACGACTTGTCT 135792
QY 309 ----- 309
Db 135791 TGGCCACACTGTGCTGGGCTGGCAAGATATGTGGAGCTGAACTTTAAAGTAGCTTCC 135732
QY 309 ----- 309
Db 135721 TCCCTCCACAGAGCCTGGGGCTGGGATCTGGTGGGATGGGGCCCTCTGGCTGCAGG 135672
QY 309 ----- 309
Db 135671 TGGCAGATGACAGAGAGAGTATGAGAGAGAGATGCTGAGAACGAAGGACTAT 135612
QY 309 ----- 309
Db 135611 CAGGAATGGCTGTGGGGAGGCTGCTGCCATGGAGGCTGTGTGTGATGACAGGTTGGCAG 135552

QY 309 ----- 309
Db 135551 AGGTCTGAAAGCGGTGAGAAATTCOCATTAATCCCTTGGCTCCAGTGAAGTAAGCAAG 135492
QY 310 ----- ArgPheSerGlyLeuLeuHisGlySerPro 319
Db 135491 GCCTACCCCTCTTCTCTCTCCCTACACACAGGTTCTGGGCTGTACATGAAATGCC 135432
QY 320 LysThrThrProSerAlaProAla ----- 327
Db 135431 AAGACACACCTTCTGCTCTGCTGTGTAATTAAGGGGGAAGGCTAATGGAGAGGGGAC 135372
QY 327 ----- 327
Db 135371 ATGTCTCCACCACTGGGGATGAGAGACCTCTATAGCATTTATGGAACACTAATGTC 135312
QY 327 ----- 327
Db 135311 CCAGAGGAAAAGTATTCTGTCCAGAGATGTGTGATGGATTAAAGCCCAACCCA 135252
QY 328 ----- -AlaSerThrGlnThrLeuThrProThr 336
Db 135251 GCAGTCTCAGATTGGGCTCATCTCTCTCAGCTTCCACAGAGACTGTACTCCACC 135192
QY 337 ProGlnuTrpAsnGlnLeuValTyrAlaSerIleGlnValGlnAlaThrGlnGlyAsnLeu 356
Db 135191 CTTGAGGGGATGATGTTGTGTCTACGCATCCATGCAAGTGCAGGCGACCCAGGAACCTT 135132
QY 357 AsnSerSerAlaGlnAspTyrAlaGlnAlaLeuTyrAspTyrThrAla ----- 371
Db 135131 AACTCATACACCCAGAGACTACCGGGCACTACAGCTACACCCG -ACAGTGAGGGCTTC 135073
QY 371 ----- 371
Db 135072 CCCTGTGCTGGAGACCCCTGGGGCTGTGAGCACTTAACATAGCACCAGAGCTGAAT 135013
QY 371 ----- 371
Db 135012 AGTTCAGTGACATCTGCTGGACCAAGATCTGCTGTGGATTAGATCAGGCCCTTGTGCACA 134953
QY 371 ----- 371
Db 134952 TGACCAAAAATGATGCTCTTGAAGTCCCTGGAATATTCTGACATCTTGGAAGAAG 134893
QY 371 ----- 371
Db 134892 TGGGACAGGGCACCCACAGACATCTGTCTTGAGATGACTGAATCCACACAAGCCCT 134833
QY 371 ----- 371
Db 134832 GGGCTTAAGGGGGTGAAGGGTGAAGCTTGAAGGAGGCCCTGAAGCTTGGGCTAATACACC 134773
QY 371 ----- 371
Db 134772 ACACATATAGCAGACATGTAGGCAAGCTAGATGGGCTGTACTGAGGTACATTCAGGG 134713
QY 371 ----- 371
Db 134712 GTACAGACCTGGACAGTAAACAGGGTACTCAGTAGCAGTAGATTTCTGTGATTGTCT 134653
QY 371 ----- 371
Db 134652 AGAGCTTGCCCTTACGTTGTGACACAAGCACTTGTGAGAGCTGCAGTGTGGTGGACACTGA 134593
QY 371 ----- 371
Db 134592 GAGAGTCCAGGATGGAGGTCTCCATGACAACCTGAGTTGTTAGAGGCCAAGTGCCT 134533
QY 371 ----- 371
Db 134532 TTGGGGAACCCCAACAGCCTGTCTATCTGCAAAAGAACCACTCTGAGAGCCTTAAGCATGA 134473
QY 371 ----- 371

DB 134472 CTCACAGACGCTGTCGCCCTGGAGACAAACATGACCTGCCAAGTGGGGGCCGAGAC 134413
 QY 372 -----Glnasn 373
 DB 134412 ACAGACACCCGCTGCTGGCTCCCTCCATCTACACACTTCCATTTCTTCCGACGAT 134353
 QY 374 SeraspGluleuAspIleSerAlaGlyAspIleAlaValIleLeuGluGlyLysP 393
 DB 134352 TCTGATGACCTGACATTTCCCGGAGACATCTCGCGCTCATCTCTGGAAGGAGGAT 134293
 QY 394 GYTPTPTPTThValGluArgAsnGlyGluArgGlyPheValProGlySerTyrLeuGlu 413
 DB 134292 GGCTGGTGGACTGTGGAGCGAAGCAGCACTGGCTTGTCTCGGTGCTGACTTGAG 134233
 QY 414 Lysleu 415
 DB 134232 AAGCTC 134227

RESULT 11
 IROEST015 562 bp mRNA linear PRI 19-OCT-1999
 LOCUS Homo sapiens EST from clone 123055, 5' end.
 DEFINITION AL110110
 ACCESSION AL110110.1 GI:5791558
 VERSION EST - expressed sequence tag.
 KEYWORDS Homo sapiens.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 562)
 Aufray,C., Ansoorge,W., Ballabio,A., Estivill,X., Gibson,K.,
 Lehnach,H., Poustka,A. and Lundberg,J.
 The European IMAGE consortium for integrated Molecular analysis of
 human gene transcripts
 Unpublished
 2 (bases 1 to 562)
 Carim,L., Estivill,X., Sunoy,L. and Escarceller,M.
 Direct Submission
 Submitted (25-AUG-1999) Dept. Genetica Molecular, Institut de
 Recerca Oncologica (IRO), Hospital Duran I Reyalas, Antoria de
 Castelldefels Km 2,7 L'Hospitalat de Llobregat, 08907 Barcelona,
 Catalunya, SPAIN. Tel: ++34-93-260-7775 Fax: ++34-93-260-7776 WWW
 site: http://www.iro.es e-mail enquiries: lsunoy@iro.es,
 mescarceller@iro.es
 EURO-IMAGE Consortium Contact: Aufray C
 CNRS UPR 420 - Genetique Moleculaire et Biologie du Developement
 IFR 1221 - Rue Guy Moquet 19, Batiment G - BP 8
 94801 Villejuif Cedex, FRANCE
 Tel: ++33-1-49 58 34 98
 Fax: ++33-1-49 58 35 09
 e-mail: aufray@infobiogen.fr
 This clone is available royalty-free through IMAGE Consortium
 distributors.
 IMPORTANT: This represents single read sequence of this IMAGE CDNA
 clone as supplied from the European distribution centers.
 Location/Qualifiers
 1..562
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone IMAGE CDNA clone 123055, 5' end"
 /clone_11b="Soares fetal liver spleen INFLS"
 /note="Matches EST Acc. No. R00125"

BASE COUNT 125 a 174 c 170 g 91 t 2 others
 ORIGIN

Alignment Scores:
 Pred. No.: 4.85e-29 Length: 562
 Score: 503.50 Matches: 105
 Percent Similarity: 69.75% Conservative: 8
 Best Local Similarity: 64.81% Mismatches: 18
 Query Match: 23.40% Indels: 32
 DB: 9 Gaps: 2

US-09-068-377c-1 (1-415) x IROEST015 (1-562)
 QY 255 ThrLeuGluGlyCysAspValGluGlyAspIleAsnGlyPheIleGlnSerLysSerTh 274
 DB 3 ACGCTGGAAGGCTGTCAGCATAGACCCGACATCCAGCTTTCATCCAGCCAAAGACG 62
 QY 275 GYAATGAGupProProlaProValProTyrGlnsnTyrTyrAspArgGluValThPro 294
 DB 63 GGCACAGAGCCCGCG-----79
 QY 295 LeuIleGlySerProSerIleGlnProSerCysGlyValIleLysArgPheSerGlyLeu 314
 DB 80 -----TTCCTGCAATG 91
 QY 315 LeuHisGlySerProLysThThPro---SerAlaProAlaAlaSerThGluThLeu 333
 DB 92 CTGCACGGAAGTCCCAAGACCACTTCCTGTCGCTGCTGCACACAGACCTG 151
 QY 334 ThrProThrProGluArgAsnGluLeuValTyrAlaSerIleGluValGlnAlaThrGln 353
 DB 152 ACCCCACCCCGAGCGGATGAGGCTGTCTACACAGCCATTCGACAGTGCAGACATTCAG 211
 QY 354 GlysAsnLeuAsnSerSerAlaGlnAspTyrArgAlaLeuTyrAspTyrThAlaGlnAsn 373
 DB 212 GGAAACCCGCGCTACACAGCCAGAGTACCGGGGCTCTACGATTATACAGCCGACAGAC 271
 QY 374 SeraspGluleuAspIleSerAlaGlyAspIleAlaValIleLeuGluGlyLysP 393
 DB 272 CCAGATGAGCTGGACCTCCACGAGACATCTCGAGAGTATCTGGAAGGAGGAGAT 331
 QY 394 GYTPTPTPTThValGluArgAsnGlyGluArgGlyPheValProGlySerTyrLeuGlu 413
 DB 332 GGCTGCTGACTGTGGAGAGGAGACGAGCGAGCTGCTTCCTGCTTCTACTGAG 391
 QY 414 Lysleu 415
 DB 392 AAGCTT 397

RESULT 12
 LOCUS AC123432/C
 DEFINITION Rattus norvegicus clone CH230-430H3, *** SEQUENCING IN PROGRESS
 AC123432
 AC123432.1 GI:21240367
 AC123432
 HTG: HTGS_PHASE1
 Rattus norvegicus.
 Rattus norvegicus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 173635)
 Muzny,D.M., Adams,C., Adio-Obiola,B., All-oman,F.R., Allen,C.,
 Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T.,
 Barbarella,J., Benton,J., Bimege,K., Blankenburg,K., Bonin,D.,
 Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlik,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hogues,M., Hollway,C., Hollins,B.,
 Homs,J.F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovari,C.,

[illegible]

QY	303	ProSerCysGlyValIleLeys-----	309
Db	165265	ACCTCTGTGGTGTATAAAGAGGTAGTACTGATGCACTGCATATAGCCAGGACAGC	165206
QY	309	-----	309
Db	165205	CGCTGGGGCCAGGCTCTCTCGCCACAACTTCACACACTCATGGCTGAAACAGCTCTCTG	165146
QY	309	-----	309
Db	165145	GGTCTGTGCCCCGTTCAAGTGTGGCACTCCATTGAGATATTCCAGATGCATCACCATC	165086
QY	309	-----	309
Db	165085	AGGGATCCCTGGTCCAGTACTAATTACTACACATCCACTTACTACCAAGTACGATG	165026
QY	309	-----	309
Db	165025	GAGATCCTGGGCAAGGCCCACTGTGCTTCTGTCCTGTTTGGCCTCATATGGCA	164966
QY	309	-----	309
Db	164965	GGGTGAACCTCCAGAGGGCTTATCAGTGGCCAAAGCCAGACATTGGATCCACG	164906
QY	309	-----	309
Db	164905	GCCGTTCAGAGGCCATCTCCCTGGCCTGGCTCAGTAAGTGTCAAGGCCATGGTCTGT	164846
QY	309	-----	309
Db	164845	GCCCTGTGGGTGACCTGGCCTAGTCAGTGGGGCCAAAGGAAGAACTTGCTTGCA	164786
QY	309	-----	309
Db	164785	CAGACTGGCTGGCCCTGGCAAGATATGTGGAGCTGAACCTTAAGTAGTGTGGGCTG	164726
QY	309	-----	309
Db	164725	GGAAATCAAGTGGGGGTGGGGCCCTGCTGGCTGCAGGTGCTGAGATGACATGGG	164666
QY	309	-----	309
Db	164665	AGGATGCTGCTGAGAACCAAGGGACTATTGAAGGACTATTGGGAATGGCTGTGGAGG	164606
QY	309	-----	309
Db	164605	CCTGTGCCACAGAGGCTGTGTGGCAGCAGGCAAGGAGTCTTGAAAGCATGAGAA	164546
QY	309	-----	309
Db	164545	TTCCCACTTACCCCTTTGTTCCAAATGAGTAAGGCAAGGCTTACCCTCTCTCTCC	164486
QY	310	-----ArgPheSerGlyLeuLeuHisGlySerProLysThrThrProSer---AlaPr	326
Db	164485	TACCTCCAGGTCTCTGTGGGTGCTCATATGGAAGTCCCAAGACACACCTTCTCAAGCTCC	164426
QY	326	Ala-----	327
Db	164425	TGCCGGTAATATGACAGAGGGGTGGGGGCAATGGGACAGGAGCATGGAGCTGTCTTC	164366
QY	327	-----	327
Db	164365	CACCACTGGGAGAGAGAGACTCCCAAAAGGCAAGGCCATTCTGTCCAAAGAGTGTGTG	164306
QY	327	-----	327
Db	164305	AAGGAAGACTAAGTTCACCCCGGCACTTCACAGGACACTGTCTCAGAGGGGGCA	164246
QY	328	-----AlaSerThrGlyThrThrLeuThrProThrProGlyLys	339
Db	164245	GGATTCACCACTAATCTGCTCTCAAGCTTCACAGACTGTGGGCTCCACCCTGTGAGCA	164186

OY		33	gysnsluleuValITyrAlaSerIleGluValGlnAlaThrGlnGlyAsnLeuAsnSerSe	359	
Db		164185	GAAAGATTGGCTATGCATCCATAAGTGAACGCCAACCAGGGAACTTAACACAGC	164126	
OY		359	rLagInAspTyrArgAlaLeuTyrAspTyrThraIa	371	
Db		164125	AACCCAGACTACCGGGCCCTTAGTACTACACGGC-ACAAGTGAGGCATCTTCCCA	164067	
OY		371	-----	371	
Db		164066	ACCCTGGGACATGCCGTGGGGGGAGGCGTAGCACACAGCAACACAGGACTGSAGA	164007	
OY		371	-----	371	
Db		164006	GCTCCATGAAATGCTGTAGAGCCAGCATCTGTGAACCTGCACGTGTGAGTCCAAAG	163947	
OY		371	-----	371	
Db		163946	GCTGCAGCTCAGGGTTGTTGTGATGATTAATGGTTBAGGCTCCGGCTTGTTCOA	163887	
OY		371	-----	371	
Db		163886	GCTGTGGCCACAGAAAGACAAATGCTCAAAAATCGAGCCATGGAGCTGCCCTTAATTT	163827	
OY		371	-----	371	
Db		163826	TTCGTGTCCTTGGAGAAAGGTGATCAGGAAATCACAGAGCATCTGACTAGAGAGAT	163767	
OY		371	-----	371	
Db		163766	ACGGATCGAGTGAAGTGAAGCGGTGGCCGGTATGACCTTGAAGGACAGCCCTCACTAG	163707	
OY		371	-----	371	
Db		163706	GCCATATATCCACACACATAGGACGGCTGATGGGGCTGCACCTGAGGTACATTACGG	163647	
OY		371	-----	371	
Db		163646	GTTACAGACCTGACAGTAATGGGATACTCAGTAGCAGTAGATGGCTGTGTATTCGTCT	163587	
OY		371	-----	371	
Db		163586	AGGACAGCTTGCCCTTGGGGTGTGGCCGCCCTTTTGAAGAGCTGCAGTGTGAGCAGTG	163527	
OY		371	-----	371	
Db		163526	AGAGAAGCCAGGATAGAGAGGTCTCCATATTACCTTAGACCGCGTTAGAGGAAACAGCAC	163467	
OY		371	-----	371	
Db		163466	GGCTGTCTATCTTAAGGAACACTACCCGAGGCTTGACATGATTCACGCCAGCTGTCA	163407	
OY		371	-----	371	
Db		163406	CCCCGGACACAAACATGCCAATGCTGCAAGGTGGGGGCTGCCACAGACAGACCCCTGG	163347	
OY		371	-----	371	
Db		163346	TCCCTGGTGGCTTCTCCATTTCTACACACACAGGCCCTTTCTTGCCACAGATTCTGATAGCT	163287	
OY		371	uAspIleSerAlaGlyAspIleLeuAlaValIleLeuGlnGlyGluAspGlyTrpTrpTh	397	
Db		163286	GGACATTCTGTGGGAGACATCTGGCAGTCACTCCGAAAGGAGAGATGGCTGGTGGAG	163227	
OY		397	rValGlnArgAsnGlyGlnArgGlyPheValProGlySerTyrLeuGlnLysLeu	415	
Db		163226	TGTGGAACGAACGAGAACAGTGGCTTGTCTCCCTGGTTCGTACTGGAGAAGCTC	163172	
RESULT 13					
AKO23100		2591 bp	mRNA	linear	PRI 01-ANG-2002
LOCUS					
DEFINITION			Hom sapiens cDNA FLJ13038 fis. clone NT2RP3001272, weakly similar		
			To Mus musculus mRNA for macrophage		

actin-associated-tyrosine-phosphorylated protein.
 AK023100
 VERSION AK023100.1 GI:10434864
 KEYWORDS Oligo capping; fls (full insert sequence).
 SOURCE Homo sapiens teratocarcinoma cell_line:NT2 CDNA to mRNA,
 clone_lib:NT2RP3 clone:NT2RP3001272.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS 1
 Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
 Nishikawa,T., Nagai,K., Sugano,S., Shitatori,A., Sudo,H.,
 Wagaatsuma,M., Hosoliri,T., Kaku,Y., Kodaira,H., Kondo,H.,
 Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
 Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
 Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
 Nakamura,Y., Nagahari,K., Masuo,Y., Niomiya,K. and Iwayanagi,T.
 NEDO human CDNA sequencing project
 TITLE Unpublished
 JOURNAL 2 (bases 1 to 2591)
 REFERENCE Isogai,T. and Otsuki,T.
 AUTHORS Direct Submission
 JOURNAL Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,
 Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 252-0812, Japan
 (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
 NEDO human CDNA sequencing project supported by Ministry of
 International Trade and Industry of Japan; CDNA full insert
 sequencing: Research Association for Biotechnology; CDNA library
 construction, 5'- & 3'-end one pass sequencing and clone selection:
 Helix Research Institute (supported by Japan Key Technology Center
 and Department of Virology, Institute of Medical Science,
 University of Tokyo).
 FEATURES
 source location/Qualifiers
 1..2591
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="NT2RP3001272"
 /cell_line="NT2"
 /cell_type="teratocarcinoma"
 /clone_lib="NT2RP3"
 /note="Cloning vector: pME18SFL3-mRNA from NT2 neuronal
 precursor cells after 2-weeks retinoic acid (RA)
 induction."
 31..669
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAB14404.1"
 /db_xref="GI:10434864"
 /translation="MDAIIKOKSIOFKKTMADKKNYEOKCRKDEAEQAVSANAIV
 PPOOKLIPYKIAKSTAVDSKAYMLHICGDKYREBQSHIKACEAFKQECERI
 NFERNAIMIHVNOISQCCYTSDEMTQVAKSLKEMCSICQDITFYVNRKTQIPAPPI
 MHEFYSSOKNAVPAKATGAPLARGPLPIPKSSPDNPYSLVDYSLIYQ"
 BASE COUNT 753 a 548 c 547 g 743 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,02e-24 Length: 2591
 Score: 452.50 Matches: 90
 Percent Similarity: 60.75% Conservative: 40
 Best Local Similarity: 42.06% Mismatches: 79
 Query Match: 21.03% Indels: 5
 Gaps: 2
 DB: 9
 US-09-068-377c-1 (1-415) x AK023100 (1-2591)
 QY 113 LysGluGlnArgLysLysTyrGlnAlaIleMetAspArgValGlnLysSerLysLeuSer 132
 Db 1 AAACACCAACGAAAAAAGACAGAGCTCATATGATGCTATCATCCATAAACAAGAGCTTA 60
 QY 133 LeuTyrLysLysThreMetGluSerLysLysAlaTyrAspGlnLysCysArgAspAlaAsp 152
 Db 61 CAATTCAAGAAACCATGATGATCAAGAAAGAAATATGAGCAGAAATGCCGGACAAAGAT 120

QY 153 AspAlaGluGlnAlaPheGluArgValSerAlaAsnGlyHisGlnGlnValGluLys 172
 Db 121 GAGCAGACAGAGCGCCGACCGAGTGGCCACACCTGGTGAACCCGAACGAACAAGAAAG 180
 QY 173 SerGlnAsnLysAlaLysGlnCysLysGlnSerAlaThrGlnAlaGluArgValTyrArg 192
 Db 181 CTTTGTGTAACCTGCGCACTTCAAGACCCGATGACAGGACGACGACAAAGCATCATG 240
 QY 193 GlnAsnIleGluGlnLysGlnArgValArgThrGluTyrGlnGlnGlnHisArgThrThr 212
 Db 241 CTCGACATCGGACACCGCTCGATAGTACGCCGAGAGGAGGAGGAGTACGACATCAAGCC 300
 QY 213 CysGlnAlaPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 232
 Db 301 TGGAGGACATTTAGGCTCAGAAATGAGACGAATAACTCTTCGCGAATGATGATGG 360
 QY 233 ValHisCysAsnGlnLysSerMetGlnCysValLysAspAspGluLysArgGlnGlnVal 252
 Db 361 TTACATGTGATGACGCTGTACACACATGTGTACACGATGATGAATGTACGAACAGTC 420
 QY 253 ArgLeuThrLeuGlnGlnCysAspValGlnGlnLysPheIleAsnGlyPheIleGlnSerLys 272
 Db 421 CGAAGAGATTTAGAAATGTGACAGCATTCAGAGGAGCATTAATCTTGTGAATCAAGCC 480
 QY 273 SerThrGlyArgGluProProAlaProValProTyrGlnAsnTyrTyrAspArgGlnVal 292
 Db 481 AAACGTGACAGATTCACACGACGACCATCATGACGAAGAATTTCTACTCTCCCGAAG 540
 QY 293 ThrProLeuIleGlySerProSerIleGlnProSerCysGlyValIleLysArgPheSer 312
 Db 541 AATGACAGTCCAGCAGCAAGAGGTACAGGCCCTAAC-----TTGCAAGAGAGCA 591
 QY 313 GlyLeuLeuHisGlySerProLysThrThrProSerAlaPro 326
 Db 592 CCCCTC-----CAATTCTTAAGCTCCACGACAGATGATGCC 627
 RESULT 14
 AC117072/c
 LOCUS AC117072
 DEFINITION Dictyostelium discoideum chromosome 2 map 3426827-3542314 strain
 AX4,*** SEQUENCING IN PROGRESS ***, in ordered pieces.
 AC117072.1 GI:20066239
 VERSION AC117072.1
 KEYWORDS HTG; HTGS; PHASE2.
 SOURCE Dictyostelium discoideum.
 ORGANISM Dictyostelium discoideum
 Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
 REFERENCE 1 (bases 1 to 115489)
 Gloeckner,G., Eichinger,L., Szafarski,K., Pachebat,J., Dear,P.,
 Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Knappf,K.,
 Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
 Noegel,A.A.
 Sequence and Analysis of Chromosome 2 of Dictyostelium
 TITLE Unpublished
 JOURNAL The Dictyostelium Genome Sequencing Consortium
 REFERENCE 2 (bases 1 to 115489)
 AUTHORS Baumgart,C.
 TITLE Direct Submission
 JOURNAL Submitted (06-APR-2002) Genome Analysis, Institute of Molecular
 Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
 REFERENCE 3 (bases 1 to 115489)
 AUTHORS Baumgart,C.
 TITLE Direct Submission
 JOURNAL Submitted (18-MAY-2002) Genome Analysis, Institute of Molecular
 Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
 CDS predictions from Genetd may contain errors. Further information
 is available from IMB Jena, Department of Genome Analysis
 (http://genome.imb-jena.de/dictyostelium/) and the University Cologne, Institute for Biochemistry I
 (http://www.uni-koeln.de/dictyostelium/project.shtml)
 Funding Agency: Deutsche Forschungsgemeinschaft (DFG).
 * NOTE: This is a 'working draft' sequence.

CDS
complement(join(24934..26525,26681..26777))
/note="ORF_ID:dd_01651"
/codon_start=1
/product="Sequence 3 from Patent WO075321. 6/101"
/protein_id="AA03154.1"
/db_xref="GI:20976565"
/translation="MNOVAESINKSLKSTADYTPKGVHTELESSTYIDHEPSSN
AKIPNPAAVAVLEQALAEHGSATSTGALVTGSGVTKGSPKRIKRVKRPSSDDI
WMGPVNIAMDLSFMINREPRADYNTQEKLYVIDGAYGMDPKRIKRVKRPSSDDI
LEHNMILIRPARELIRNRPEDYITACOPFAPARYKSSSSSIDFARKEMTI
LGTOYAGSKKILITMYLPMKGVLPKHSNQRNNGDTLFFGLSGTKTLLA
DINRELSDDEHWTDCENI EGGCAKCIDLSKEKEPIFDAIKFGAVLENNYNE
TSRKVDVNDVSTENTRCAPYELHLPNKPFAIAGHPNIIIMLCDAFGILPVSRLD
ANQVMYHPIQGYTAQVAGTEVEPTATFSSCYGEPYIWHPTKMYEMLAISOLHKS
ARAWLINTGWTGSHGVSRILKLAATRAIIDAISGELEKIPTKMDYVFGQVNSCP
GVPSILMPINGMDKEKYSMTMKLAKFTENKRRKODKASPELVAVGPILPQ"
join(28969..29084,29176..29317,29453..30610)
/note="ORF_ID:dd_00794"
/codon_start=1

Alignment Scores:
Pred. No.: 2,48e-12 Length: 115489
Score: 317.50 Matches: 105
Percent Similarity: 41.14% Conservative: 97
Best Local Similarity: 21.38% Mismatches: 182
Query Match: 14.75% Indels: 107
Gaps: 15

US-09-068-377c-1 (1-415) x AC117072 (1-115489)

QY 1 MetMetLagInLeuGInPheArGAsPaLaPheTrPcysArGAsPheThrAlaHisThr 20
DB 23967 ATGTCACACACAGACAAATTTTCAGAAAGATCTTTGGGATGAA----- 23926
QY 21 G1YtYrG1uValLeuLeuGInArGLeuLeuAspL1yArG1yMetCysLysAspValGlu 40
DB 23925 ---TTGATACAGTGTGTTAAAAAGTGAATATGCTAAATTTTCACACACAAATTTAGT 23869
QY 41 G1uLeuLeuArGInArGAlaGInAlaG1uG1uArG1yG1yG1uLeuValG1uLeu 60
DB 23868 AAAATTTCTAAGTAAACAAACAATTTGAATGACCCATGCAAGAGTTAGTAAAGTTA 23809
QY 61 Ala-----ArgLysAlaG1yG1uInThrG1uMetAsnSerLeuArgThrSerPheAsp 78
DB 23808 TGTAAAGATTAATCATTTGCACACAGAGTGTGAATGGGACATTTGAGAGATTCATTCA 23749
QY 79 SerLeuLysG1uG1uInThrG1uAsnValG1ySerAlaHis1LeuLeuAlaLeuAlaLeu 98
DB 23748 TGTATATCGTGAACAAATTTGATGAGCATTCATGAGAGAAATTTCAAAATGATTA 23689
QY 99 ArgLysG1uLeuArg---SerLeuG1uG1uPheArG1uArG1uG1uLysG1uG1uArgLys 117
DB 23688 GAGAAATTTAGTACCATTTGATTCGATGATTTTGGAGAGAGATGAAGAAACAAAGAAA 23629
QY 118 LysTyrG1uAla1LeuMetAspArGValG1uLysSerLysLeuSerLeuTyrLysLysThr 137
DB 23628 GCATTGATTCGCAATGTTGTAAGAGTGTACAAAGATCTTAACACTCCGAGCAATGAG 23569
QY 138 MetG1uSerLysLysAlaTyrTrsPglInLysCysArGAsPaLaAsPaLaG1uG1uAla 157
DB 23568 TCAAAAGCAAAAGCAAACTATGAAAAAGTTAAAGAGAAACAAGAGAGAGCCCAATGAGAT 23509
QY 158 PheG1uArGValSerAlaAsnG1yHisG1uLysG1uValG1uLysSerG1uAsnLysAla 177
DB 23508 CTCCTCAAAA-----CAACTCCAGGCCGCCCAAAAGCAACA---AAGGCC 23470
QY 178 LysG1uG1yLysG1uSerAlaThrG1uAla-----G1uArGValTyrArG1u 193
DB 23469 AGAAAGACATTGGATTCGCAACTAAACCGCTGACAAAGGATTAATGACATTAAGGAGAG 23410
QY 194 Asn1LeuG1uInLeuG1uArG1uArG1uThrG1uInG1uG1uHisArG1uThrCys 213
DB 23409 AGTGTGAATGTTTTCACAAACAATTCACGAATTTCTATCATGAAGAGATCCCAAGATTT 23350

QY 214 G1uAlaPheG1uLeuG1uInPheAspArgLeuThr1LeuLeuArgAsnAlaLeuTyrVal 233
DB 23349 ---TTAGATGATCTTCAAAAGTTTGAAGTTGAAGAAATGATTAATCAAAAGATGGTTA 23293
QY 234 HisCysAsnG1uLeuSerMetGlnCysValLysAspArgLeuTyr----- 249
DB 23292 -----ATGGAAGTTATCATCTCCAAATGAATTCGATACACACCCGCTGT 23251
QY 250 -----G1uG1uValArGLeuThrLeuG1uG1uLysAspValG1uG1uAsp1Le 265
DB 23250 ATCATTAATATGAACAAATATTAAGAAAGTATGAATCTATGCAATGAGACAGCTGATCT 23191
QY 266 AsnG1yPhe1LeuG1uSerLysSerThrG1yArG1uProProAlaProValProTyrGln 285
DB 23190 CAATATATATCTTGTGAACAAATGAGAGTGTGCTCAAAACCCACAGACCTCAATATGA 23131
QY 286 AsnTyr----- 287
DB 23130 CTTATTCATATCTGTGTGTGTTTGCATATGATTAATGATTAATGATTAAT 23071
QY 288 TyrAspArgG1uValThrProLeu1LeuG1ySerProSer1LeuG1uProSerCysG1uVal 307
DB 23070 ATTAAGTCGTAAATCTGTGATTAATGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 23011
QY 308 IleLys-----ArgPheSerG1yLeuLeuHisG1ySerPro----- 319
DB 23010 ATTTCTTCCACACACACACACACAAATTCAAATATGATCATCAAAATCTCCACACACCA 22951
QY 319 ----- 319
DB 22950 AATATATATATACAGCATACATACATCAACAAACAAACACACACACAAATATAGT 22891
QY 320 -----LysThrThrProSerAlaProAla1AsnSerThrG1uThrLeuThrProThrPro 337
DB 22890 AATGTAATGACAAATCTCCACACACACCA-----CAACACACACCA 22852
QY 338 G1uArGAsnG1uLeu-----ValTyr 344
DB 22851 CAACACACAAACAATTTCTCCACACACAACTGATTAATTAATCCACACACACCAATTTCT 22792
QY 345 AlaSer1LeuValG1uAlaThrG1uG1uLysLeuAsnSerSerAlaG1uAspTyrArg 364
DB 22791 TTAATTAATAAATGATAGTATAGTATAGTAAATAGTAAATAGTAAATAGTAAATAGTAA 22732
QY 365 AlaLeuTyrAspTyrThrAlaGlnAsnSerAspLeuAsp1LeuSerAlaG1yAsp1Le 384
DB 22731 GCATTTATGATTAATTAATGCAACAGAGAGATGAATTCATTTAAAGCAATGCTTTA 22672
QY 385 LeuAlaVal1LeuLeuG1uG1yG1uAspG1yThrTrp-----ThrValG1uArGAsn 401
DB 22671 ATTAAGTTGTTTATACGAGCAAGATGTTGTGGCAAGGTATGTAATTTGAGAGAGT 22612
QY 402 GlyGlnArG1yPheValProG1ySerTyrLeu 412
DB 22611 GATCGTATTTGTTTCCCAAGTAATTTATTT 22579

RESULT 15
BC008216
LOCUS BC008216
DEFINITION Homo sapiens, similar to proline-serine-threonine
phosphatase-interacting protein 2, clone MGC:9914 IMAGE:3871158,
mRNA, complete cds.
ACCESSION BC008216
VERSION BC008216.1 GI:14198308
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1461)
AUTHORS Strausberg R.
TITLE Direct Submission

JOURNAL

Submitted (22-MAY-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: villalobebcm.tmc.edu

Villalob, D.K., Luna, R.A., Hale, S.M., Huylk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Series: IRK Plate: 13 Row: K Column: 8
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, similarity but not identity to protein.

FEATURES

SOURCE

1..1461

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="MGC:9914 IMAGE:3871158"

/tissue_type="Eye, retinoblastoma"

/clone_id="NIH_MGC_67"

/lab_host="DH10B"

/note="vector: PCMV-SPORT6"

79..612

/codon_start=1

/product="Similar to proline-serine-threonine phosphatase-interacting protein 2"

/protein_id="AAH08216.1"

/db_xref="GI:14198309"

/translation="MTRSLFKGNFWSADILSTIGYDNIIOHNLNCRKCKEEDFLKE RALIERGKDLNLSRKPKCGSEINTLRALVEFKQVDNVAOCCHIOAQLRREA RKMEEFREKOKLQRRKMEHSVTOAGACQWHDGSLDALPRGFMPFCLSLPSSWNYRL PPPRLAEPBNDHGVA"

PPPRLAEPBNDHGVA"

BASE COUNT

432 a 327 c 369 g 333 t

ORIGIN

Alignment Scores:

Pred. No.: 7,96e-14 Length: 1461
Score: 305.50 Matches: 59
Percent Similarity: 70.25% Conservative: 26
Best Local Similarity: 48.76% Mismatches: 35
Query Match: 14.20% Indels: 1
DB: 9 Gaps: 1

US-09-068-377c-1 (1-415) x BC008216 (1-1461)

QY 2 MetAlaGlnLeuGlnPheArgAspAlaPheTrpCysArgAspPheThrAlaHisThrGly 21
DB 79 ATGAGCGCGTCACTGTTCAAGGAAACTTTGGAGTCAGACATCCTCAGACACATCGGC 138
QY 22 TyrGluValLeuLeuGlnArgGluLeuAspGlyArgGlyMetCysLysAspValGluGlu 41
DB 139 TATGACAACTATTCACATCTGAACATGGCCGCAAGAACCTCAAGAGTTTGAAGAC 198
QY 42 LeuLeuArgGlnArgAlaGlnAlaGlnGluArgTyrGlyLysGluLeuValGlnIleAla 61
DB 199 TTTCATAAAGAAGGACACATGTGAAGAGGATATGCAAGATCTGCTCAACCTCTCT 258
QY 62 ArgLysAlaGly---GlyGlnThrGluMetAsnSerLeuArgThrSerPheAspSerLeu 80
DB 259 AGGAAGAAGCCGTGTGACAGTCTGAAATCAACACACCTGAAGCGGCGCTTGAAATCTTC 318

QY 81 LysGlnGlnThrGluAsnValGlySerAlaHisIleGlnLeuAlaLeuAlaLeuArgGlu 100
DB 319 AACGACCACTAGACATGATGCGCACATGTCACATTCAGCTTGCACAGACTTTAAGAGAA 378
QY 101 GlnLeuArgSerLeuGlnGluGluPheArgGluArgGlnLysGlnArgLysTyrGlu 120
DB 379 GAGGCCAGGAAGATGAGAGAAATTCAGGGAAGCAAAAGCAAAACTACAGCAAGAAAGATGAG 438
QY 121 Ala 121
DB 439 TCC 441

Search completed: August 3, 2003, 13:20:34
Job time : 3676 secs

THIS PAGE BLANK (USPTO)